Smoothed GMM for Spatial Autoregressive Quantile Models

HU Yanan[a], QU Xinhao[b], Wolfgang Karl Härdle[c], TIAN Maozai[d]

[a] *School of Business, Zhengzhou University, Zhengzhou, 450001, Henan, China*
[b] *School of Ecomomics, Xiamen Univeristy, Xiamen, 361005, Fujian, China*
[c] *School of Business and Economics, Humboldt Universitat zu Berlin, Berlin, Germany*
[d] *School of Statistics, Renmin University of China, Beijing, 100872, Beijing, China*

Abstract

In consideration of cross-sectional dependence, heterogeneity, and potential outliers, we employ a spatial autoregressive model utilizing quantile regression. The methodology offers several advantages: Comprehensive Representation: It has the capability to unveil overall properties of the conditional distribution. Quantitative Spatial Insight: The model excels in describing spatial effects across various quantile levels. Enhanced Robustness: Its increased robustness enables adaptation to a broader spectrum of spatial error structures. To address endogeneity arising from the spatial lag and the differentiability of the objective function, instrumental variables are employed in this paper. A smoothed moment condition is established to render the objective function differentiable. Furthermore, an optimal bandwidth is selected to estimate the parameter. We explicitly demonstrate the large sample properties of consistency, asymptotic normality, and efficiency for the estimators. Our proposed method exhibits superior speed and performance in finite sample scenarios, as demonstrated through simulation research. Finally, leveraging the spatial quantile regression model, we delve into the heterogeneity and spatial aggregation effects of rural labor transfer on the incidence of rural poverty.

Keywords: Spatial Autoregressive Model, Quantile Regression, Smoothed GMM

1. Introduction

In the realm of spatial data modeling, cross-sectional dependencies arise from the omission of pertinent variables, measurement errors, or spillover effects. Spatial autoregressive models (SAR) strategically introduce spatial lag terms to capture the cross-sectional dependencies between observations, a phenomenon highly scrutinized by both econometricians and statisticians. Cliff and Ord (1984) [1] systematically summarized the theoretical methodologies for parameter model estimation and inference. Anselin and Bera (1998) [2] provided a comprehensive description of the spatial autoregressive linear model. However, owing to the endogeneity of the spatial lag term, the utilization of ordinary least squares estimation leads to biased and inconsistent estimators. This issue, constituting a pivotal aspect of spatial econometric modeling, has garnered significant attention in the academic literature. Numerous works have delved into the statistical inference and application research of spatial autoregressive models (Drukker *et al.*, 2014 [3] ; Qu and Lee, 2015[4] ; Jin and Lee, 2018[5]).

It is essential to highlight that the aforementioned studies primarily adopt a mean regression perspective, emphasizing the expected information within the conditional distribution of the response variable and demonstrating sensitivity to outliers. In the context of spatial data modeling, this paper diverges by undertaking research from the standpoint of quantile regression. Our primary considerations are twofold: Firstly, spatial data typically exhibit heterogeneity. As noted by Koenker (2005)[6], the impact of a covariate on the center of the conditional distribution of the response variable can significantly differ from its effect on the tails. Consequently, concentrating solely on the conditional mean function may lead to misleading insights. By estimating conditional quantiles at various levels, a more comprehensive understanding of the relationship between covariates and response variables can be achieved. Secondly, quantile regression possesses a natural robustness to outliers associated with heavy-tailed errors. This robustness inherently shields against the influence of outliers in spatial data, providing a degree of protection during outlier identification. The spatial quantile regression model innovatively integrates quantile regression techniques into spatial data modeling, thereby broadening the research perspective of the spatial autoregressive model. This represents a nascent field within spatial econometrics, with research in this domain still being relatively limited.

Spatial quantile regression models offer numerous advantages, yet they encounter challenges in the estimation process, particularly in addressing the endogeneity of lagged factors and the non-smoothness of loss functions. Primarily, the spatial lag term, when treated as a covariate, becomes an endogenous variable, and estimators obtained through traditional quantile regression methods lack consistency. Notably, techniques such as two-stage quantile regression (Kim and Muller, 2004[7]) and instrumental variable quantile regression (Chernozhukov and Hansen, 2008[8]) have emerged as key approaches to tackle endogeneity in spatial quantile regression models. While not specifically designed for spatial econometric models, these methods have been extensively applied by researchers such as Su and Yang (2007)[9], McMillen (2012)[10], and Li and Fang (2018) [11] in estimating spatial autoregressive models using instrumental variable quantile regression. Zietz *et al.* (2008)[12] conducted research on the spatial quantile regression model based on two-stage quantile regression, and Kostov (2009)[13] compared two-stage quantile regression and instrumental variable quantile regression methods, noting superior small sample properties of the latter. Secondly, the non-differentiability of the objective function poses a substantial challenge. The quantile regression loss function, being a weighted loss, renders the objective function non-differentiable, presenting significant computational complexities and complicating statistical inference. The asymptotic covariance matrix, dependent on the overall conditional density evaluated at the true quantile, introduces challenges in calculating asymptotic confidence intervals. Additionally, the asymptotic normality of the standard quantile regression estimator relies on the Bahadur-Kiefer representation, which exhibits poor convergence. In response, Kaplan and Sun (2017) [14]and De Castro *et al.* (2019)[15]have made strides by constructing smooth moment conditions and creating differentiable objective functions for estimating parameters. These advancements contribute to overcoming the computational difficulties associated with non-differentiable objective functions in spatial quantile regression models.

This paper delves into the study of the spatial quantile regression model for spatial data, proposing a novel smoothing generalized moment estimation method. The method boasts several advantages: (1)Adaptability to Weaker Conditions: In comparison to the previous instrumental variable quantile regression by Chernozhukov and Hansen (2008) [8], the proposed method demonstrates a greater capacity to adapt to weaker conditions.(2) Derivation of Unconditional Moments: Through leveraging the constraints of the conditional quantile of the disturbance term, this method derives unconditional moments, satisfying the sufficient conditions for local identification.

(3) Computational Efficiency: In contrast to the smooth quantile regression loss function proposed by Horowitz (1998) [16], the proposed method reduces the deviation of the smooth moment function estimator by multiples, thereby offering significant computational advantages (Kaplan and Sun, 2017 [14]). (4) Enhanced Analytical Expressions: The non-smooth sample moment of quantile regression typically lacks an analytical expression. However, by assuming a smooth function, the proposed method decreases the asymptotic variance of the smooth moment function estimator. (5) Superior Estimation Performance: Simulation studies affirm that, when compared to instrumental variable quantile regression, the estimators obtained through this method exhibit greater accuracy and robustness. The method outperforms in small sample scenarios, demonstrating clear computational advantages. In summary, the proposed smoothing generalized moment estimation method not only adapts well to varying conditions but also offers analytical and computational advantages, making it a robust and efficient approach for spatial quantile regression modeling in the context of spatial data.

2. Model

In the analysis of spatial data, where cross-sectional dependence is inherent, the specification of a spatial weight matrix becomes crucial for describing spatial dependence relationships. Additionally, introducing the spatial lag item as a covariate into the model is a common practice. Whether conducting mean regression or quantile regression, due consideration must be given to the potential endogeneity arising from the inclusion of the spatial lag item.

2.1. Spatial Autoregressive Model

We consider the following data generating process:

$$y_i = \rho \sum_{j=1}^{n} w_{ij} y_j + x_i' \beta + \varepsilon_i, \ i \neq j \in \{1, 2, ..., n\} \tag{1}$$

where y_i denote the observation of the response variable for spatial unit i, and x_i represents the d_β dimensional exogenous variables of spatial unit i. The matrix $W_n = \{w_{ij}\}_{n\times n}$ is the spatial weight matrix, where $\sum_{j=1}^{n} w_{ij} y_j$ corresponds to the spatial lag term. The parameter β is a d_β dimensional coefficient associated with the exogenous variables and is subject to estimation. The coefficient ρ signifies the spatial autocorrelation. The disturbance term is denoted by ε_i.

The spatial weight matrix is usually row normalized so that the elements in each row sum to 1, meaning that each spatial cell's observation of the response variable y_i is weighted by the response variable observations of nearby cells y_j The influence of the form. If the coefficient $\rho > 0$, there is a positive spatial correlation, that is, the response variable observations are spatially clustered. If the coefficient $\rho < 0$, there is a negative spatial correlation, that is Neighboring y tend to be different. Considering the stationarity condition, it is guaranteed that $-1 < \rho < 1$.

2.2. Spatial quantile regression model

In mean regression, model (1) exclusively centers on the conditional mean function, often imposing stringent assumptions on the disturbance term and demonstrating sensitivity to outliers. To broaden the research perspective of the spatial autoregressive model and address these limitations, this paper integrates the quantile regression technique. By adopting the spatial quantile regression model (**??**), o2urequ:: study extends beyond the confines of mean regression. This innovative approach not only accounts for the conditional mean but also provides a more robust and flexible framework for modeling spatial data.

Under mean regression, the model (1) only focuses on the conditional mean function, and usually has strict assumptions on the disturbance term, and is sensitive to outliers. Expand the research perspective of the spatial autoregressive model, combined with the quantile regression technique To model spatial data, this paper studies the spatial quantile regression model

$$Q_\tau(y_i \mid x_i) = \rho(\tau) \sum_{j=1}^{n} w_{ij} y_j + x_i' \beta(\tau), \ i \neq j \in \{1, 2, ..., n\} \tag{2}$$

where τ is the quantile level, with its values ranging between 0 and 1. The covariate coefficient $\beta(\tau)$ varies as a function of the quantile level τ, elucidating the relationship between variables at different quantile levels. This dynamic characterization enables an insightful examination of how covariates influence the heterogeneity of the response variable. Similarly, the spatial autoregressive coefficient $\rho(\tau)$ is also contingent on the quantile level τ and adheres to the assumption of stationarity. To ensure the non-singularity of the martrix $I_n - \rho(\tau)W_n$, where W_n represents the spatial weight matrix and I_n is the identity matrix of order n, it is presumed that the absolute value of the eignvalues of W_n does not exceed 1. Furthermore, the parameter $rho(\tau)$ is constrained to lie within the unit cicle, thereby ensuring the non-singularity of the matrix product.

The spatial quantile regression model (2) offers several notable advantages: First, Robustness to Outliers and General Applicability: The model demonstrates resilience to outliers in spatial data and imposes no stringent assumptions on the disturbance term. This characteristic renders it well-suited for accommodating a broader range of spatial error structures. Second, Ability to Capture Heterogeneous Influences: A distinctive strength of the model lies in its capability to depict the heterogeneous influence between variables. This feature allows for a nuanced representation of relationships in spatial data, providing a richer understanding. Third, Quantification of Spatial Effects Across Quantile Levels: The model excels in capturing spatial effects at various quantile levels. This versatility allows for a comprehensive exploration of how spatial relationships evolve across different quantiles, offering a more detailed and nuanced portrayal.

3. Parameter Esitmate

While the spatial quantile regression model presents notable advantages, the estimation process entails several challenges, notably the endogeneity associated with the spatial lag term and the non-differentiability of the objective function. In addressing these challenges, this paper leverages the approach introduced by De Castro *et al.* (2019) [15]. The method involves constructing instrumental variables within the framework of general moment conditions for model (2). This strategy extends to smoothing the moment function, drawing inspiration from De Castro *et al.* (2019) [15], resulting in a refined approach. The outcome is a smooth generalized moment estimator, enhancing both computational efficiency and the accuracy of estimators. This innovative methodology contributes to improving the speed of calculations while simultaneously enhancing the precision of the estimators.

3.1. Instrumental Variable

To effectively address the endogeneity issue inherent in the model, a judicious selection of instrumental variables z_i is paramount. Drawing inspiration from the work of Liu and Lee (2010) [17], the assumption is made that x_i and z_i are strictly exogenous and exhibit correlation. This assumption provides a foundation for mitigating endogeneity concerns, and it is from this premise that the instrumental variables are selected. The choice of appropriate instrumental variables is instrumental in refining the model and enhancing its capacity to produce reliable and unbiased estimates.

$$W_n\mathbf{X}, W_n\mathbf{Z}, W_n^2\mathbf{X}, W_n^2\mathbf{Z}, \ldots \tag{3}$$

The choice of instrumental variables involves selecting linearly independent variables, with one viable option being the endogenous spatial lag item $W_n\mathbf{Y}$. In this context, $\mathbf{X}$ represents the vector of exogenous variables, $\mathbf{Y}$ is the vector representing response variables, and $\mathbf{Z}$ is the vector representing instrumental variables. To generate a robust set of instrumental variables, the combination of $\mathbf{X}$, $\mathbf{Y}$ and $W_n\mathbf{Y}$ is considered. This set, while not necessarily optimal, is employed to obtain a consistent estimate. Redundant items are subsequently identified and removed to streamline the instrumental variable set while preserving the consistency of the estimate. This pragmatic approach balances the need for consistency with the desire to eliminate unnecessary complexity in the model.

3.2. Moment Conditions

In the context of endogeneity, Chernozhukov and Hansen (2005) [18] conducted a thorough investigation into the formulation and identification of instrumental variable quantile regression models. Their work is particularly noteworthy for providing precise moment constraints tailored to the intricacies of quantile regression. This contribution enhances the understanding and application of instrumental variable techniques in the presence of endogeneity, offering valuable insights into the modeling process. Given exogenous covariates x_i and instrumental variables z_i, the spatial quantile regression model (2) has

$$P\{y_i \leq \rho(\tau) \sum_{j=1}^{n} w_{ij} y_j + x_i' \beta(\tau) \mid x_i, z_i\} = \tau, \quad i \neq j \in \{1, 2, ..., n\} \tag{4}$$

It means that τ-th quantile of random variable $y_i - \rho(\tau) \sum_{j=1}^{n} w_{ij} y_j - x_i'\beta(\tau)$ is 0, which satisfies

$$P\{y_i - \rho(\tau) \sum_{j=1}^{n} w_{ij} y_j - x_i' \beta(\tau) \leq 0 \mid x_i, z_i\} = \tau, \quad i \neq j \in \{1, 2, ..., n\} \tag{5}$$

The aforementioned formula can be expressed by conditional moment constraints as

$$E[I\{y_i - \rho(\tau)\sum_{j=1}^{n} w_{ij}y_j - x_i'\beta(\tau) \leq 0\} - \tau \mid x_i, z_i] = 0 \tag{6}$$

where $I\{\cdot\}$ is an indicative function. According to the law of repeated expectations (LIE), the unconditional moments can be obtained

$$E(z_i[I\{y_i - \rho(\tau)\sum_{j=1}^{n} w_{ij}y_j - x_i'\beta(\tau) \leq 0\} - \tau]) = 0. \tag{7}$$

For convenience, denote $\theta(\tau) \equiv (\rho(\tau), \beta(\tau)')'$, let

$$g_i^u(\theta(\tau), \tau) \equiv z_i\Big[I\Big\{y_i - \rho(\tau)\sum_{j=1}^{n} w_{ij}y_j - x_i'\beta(\tau) \leq 0\Big\} - \tau\Big] \tag{8}$$

$$M^u(\theta(\tau), \tau) \equiv E[g_i^u(\theta(\tau), \tau)] \tag{9}$$

where the superscript u indicates that it has not been smoothed. Then, the above moment condition (7) can be expressed as

$$M^u(\theta(\tau), \tau) = 0 \tag{10}$$

Thus, without smoothing, the sample moment condition of the spatial quantile regression model can be obtained

$$\hat{M}_n^u(\theta(\tau), \tau) \equiv \hat{E}[g_i^u(\theta(\tau), \tau)] = \frac{1}{n}\sum_{i=1}^{n} g_i^u(\theta(\tau), \tau) \tag{11}$$

where $\hat{E}(\cdot)$ represents the sample mean.

3.3. Smoothed GMM

To address the challenge posed by the non-differentiability of the quantile regression objective function, a smoothing technique is applied to the indicative function. Providing a more tractable and analytically manageable form for further analysis, the resulting smoothed sample moments are denoted as

$$\hat{M}_n(\theta(\tau), \tau) \equiv \frac{1}{n}\sum_{i=1}^{n} g_{ni}(\theta(\tau), \tau) \tag{12}$$

which

$$g_{ni}(\theta(\tau),\tau) \equiv z_i\Big[\widetilde{I}(\frac{\rho(\tau)\sum\limits_{j=1}^{n} w_{ij}y_j + x_i'\beta(\tau) - y_i}{h_n}) - \tau\Big] \tag{13}$$

In the equation, h_n symbolizes the bandwidth, and $\widetilde{I}(\cdot)$ denotes the smoothing function. While, for the sake of data simulation, it is permissible to opt for the smallest feasible bandwidth $h'_{\text{possibly}-\text{small}}$ to streamline computations, the necessity for the asymptotic normality of the estimator compels adherence to the condition $\sqrt{n}h_n = o(1)$ (refer to Appendix A.3). Drawing from the linear model of spatial quantile regression, this paper employs the Plug-in method, as outlined in Kaplan and Sun (2017) [14], to determine the window width, ensuring compliance with the aforementioned conditions. Additionally, according to insights from Horowitz (1998) [16]and Kaplan and Sun (2017) [14], the selection of the smoothing function is

$$\widetilde{I}(a) \equiv I\{-1 \le a \le 1\}\left[0.5 + \frac{105}{64}\left(a - \frac{5}{3}a^3 + \frac{7}{5}a^5 - \frac{3}{7}a^7\right)\right] + I\{a > 1\} \tag{14}$$

for simulation studies.

In cases of over-identification where $d_Z > d_\beta$, the sample moment conditional equations lack an exact solution, necessitating the estimation of parameters through generalized moments. This approach is adopted to address the surplus of instrumental variables (d_Z) compared to the number of parameters associated with the exogenous covariates (d_β), ensuring a robust estimation procedure under the given conditions. Let

$$\overline{G} \equiv \frac{1}{n}\sum_{i=1}^{n} \nabla_{\theta^\top} g_{ni}(\overline{\theta}(\tau),\tau) \tag{15}$$

where $\nabla_{\theta^\top}$ is the partial derivative of $\theta^\top$. $\overline{\theta}(\tau)$ is $\theta_0(\tau)$ initial continuous estimator. And record

$$\overline{\Omega} \equiv \frac{1}{n}\sum_{i=1}^{n} g_{ni}(\overline{\theta}(\tau),\tau)\ g_{ni}(\overline{\theta}(\tau),\tau)^\top \tag{16}$$

is the estimator of $\Omega = E\left[g_{ni}(\theta_0(\tau))g_{ni}(\theta_0(\tau))^\top\right]$, so the one-step estimator is

$$\hat{\theta}_{\text{MM}}(\tau) = \overline{\theta}(\tau) - (\overline{G}^\top\overline{\Omega}^{-1}\overline{G})^{-1}\overline{G}^\top\overline{\Omega}^{-1}\frac{1}{n}\sum_{i=1}^{n} g_{ni}(\overline{\theta}(\tau),\tau) \tag{17}$$

where $\overline{\theta}(\tau)$ satisfies the sample moment condition. The smooth generalized moment estimator is based on the principle of minimizing the weighted quadratic norm of the sample moment vector, which can be expressed as

$$\hat{\theta}_{\mathrm{SGMM}}(\tau) = \arg\min_{\theta\in\Theta} \hat{M}_n(\theta(\tau),\tau)^{\top}\overline{\Omega}^{-1}\hat{M}_n(\theta(\tau),\tau) \tag{18}$$

Given the potential non-convex nature of the function, the simulated annealing algorithm is employed to calculate the global minimum. To initiate the iteration, $\hat{\theta}_{\mathrm{MM}}(\tau)$ is used as the initial value. This approach ensures a comprehensive exploration of the parameter space, mitigating the risk of convergence to local minima and enhancing the algorithm's ability to identify the global minimum.

3.4. Method Advantage

Horowitz (1998)[16] pioneered the introduction of smoothing techniques in ordinary median regression, rendering the objective function differentiable for subsequent parameter estimation. Building on this foundation and drawing inspiration from Kaplan and Sun (2017) [14], this paper seeks to compare the smooth estimator derived from the general quantile regression loss function with that of the instrumental variable quantile regression generalized moment function. Emphasis is placed on elucidating the statistical properties of the latter method, highlighting its advantages from a methodological standpoint.

When $d_Z = d_X$, based on the sample moment expression (11), its first-order condition

$$\frac{\partial}{\partial\theta}\bigg|_{\theta(\tau)=\theta_0(\tau)} \frac{1}{n}\sum_{i=1}^{n}\left[\widetilde{I}\Big(\frac{\rho(\tau)\sum\limits_{j=1}^{n} w_{ij}y_j + x_i'\beta(\tau) - y_i}{h_n}\Big) - \tau\right]\Big(\rho(\tau)\sum_{j=1}^{n} w_{ij}y_j + x_i'\beta(\tau) - y_i\Big) = 0 \tag{19}$$

By extracting the constant term, the left-hand side of the equation can be transformed into a form denoted as

$$\frac{1}{n}\sum_{i=1}^{n}\left\{x_i\left[\widetilde{I}\left(\frac{-\Lambda_i(Y,X,\theta(\tau))}{h_n}\right)-\tau\right]+\left[\widetilde{I}'\left(\frac{-\Lambda_i(Y,X,\theta(\tau))}{h_n}\right)\left(\frac{x_i}{h_n}\right)\right][-\Lambda_i(Y,X,\theta(\tau))]\right\} \tag{20}$$

where $\Lambda_i(Y,X,\theta(\tau)) \equiv y_i - \rho(\tau)\sum\limits_{j=1}^{n} w_{ij}y_j - x_i'\beta(\tau)$. The correspondence between the first term and the smooth sample moment (12) becomes apparent, indicating that the smooth moment estimator can be perceived as the extraction and simplification of the smooth quantile regression loss function estimator. This observation underscores the computational advantages embedded in the proposed method.

For the smooth sample moment (12), the deviation of the smooth moment function can be calculated as follows:

$$\begin{aligned} & BIAS_{\text{SGMM}} \equiv E\left(g_{ni}(\theta_0(\tau),\tau)\right) \\ = \; & \frac{(-h_n)^r}{r!}\left[\int_{-1}^{1}\widetilde{I}'(v)v^r dv\right] E\left[f_{\Lambda|z}^{(r-1)}\left(0|z_j\right)z_j\right] + o\left(h_n^r\right) \end{aligned}$$

Similarly, the deviation of the smooth quantile regression loss function can also be calculated

$$\begin{aligned} & BIAS_{\text{SQR}} \equiv E\left(k_{ni}(\theta_0(\tau),\tau)\right) \\ = \; & (r+1)\frac{(-h_n)^r}{r!}\left[\int_{-1}^{1}\widetilde{I}'(v)v^r dv\right] E\left[f_{\Lambda|z}^{(r-1)}\left(0 \mid z_j\right)z_j\right] + o\left(h_n^r\right) \end{aligned}$$

where

$$k_{ni}(\theta_0(\tau),\tau) \equiv x_i\left[\widetilde{I}\left(-\Lambda_i/h_n\right) - \tau\right] + (1/h_n)\widetilde{I}'\left(-\Lambda_i/h_n\right)\left(-x_i\Lambda_i\right) \tag{21}$$

The expression provides a detailed representation of the loss function after smoothing. For a comprehensive understanding of the calculation process, refer to Appendix A.1. Assuming the existence of the density function $f_{\Lambda|z}(\cdot|z)$ is bounded, and r-th order derivatives can be derived, the detailed comparison reveals that the smoothing deviation of the former is reduced by $r+1$ times($r \geq 2$) compared to the latter. Under the premise of the same estimated variance, this reduction in deviation is advantageous for obtaining a more precise confidence interval. Moreover, under the conditions satisfying the higher-level assumptions, it can be demonstrated that $E[\hat{M}_n(\theta_0(\tau),\tau)] = o_p(1)$. Consequently, the asymptotic variance of the smooth moment function can be derived as follows:

$$\begin{aligned} & AVAR_{\text{SGMM}} \equiv E\left(g_{ni}(\theta_0(\tau),\tau)' g_{ni}(\theta_0(\tau),\tau)\right) \\ = \; & \tau(1-\tau)E\left(z_i' z_i\right) - h_n\left[1 - \int_{-1}^{1}\widetilde{I}^2(u)du\right] E\left[f_{\Lambda|z}\left(0 \mid z_i\right)z_i' z_i\right] + O\left(h_n^2\right) \end{aligned}$$

Since $1 - \int_{-1}^{1}\widetilde{I}^2(u)du > 0$, the conclusion drawn is that smoothing effectively diminishes the asymptotic variance of the estimator. The process of smoothing involves rendering the indicative function continuous and diversifying the variable values. This, in turn, diminishes the disparities between the values of each function, leading to a reduction in the asymptotic variance of the estimator. This observation is intuitively understandable as the smoothing method fosters continuity and diversity, contributing to the enhanced precision of the estimator.

In summary, in contrast to the unsmoothed moment function estimator, the smoothed estimator exhibits an enhanced asymptotic variance. Furthermore, compared with the smoothed quantile regression loss function, the deviation of the estimator undergoes a multiple reduction, highlighting pronounced computational advantages. This underscores the efficacy of the smoothing technique in refining the statistical properties and computational efficiency of the estimator.

4. Large Sample Properties

To establish the consistency, asymptotic normality, and asymptotic validity of the estimated parameters, this paper introduces the following assumptions. Each assumption corresponds to a specific facet of the argument and is sequentially addressed to construct a comprehensive justification for the statistical properties of the estimators.

1. For any observation in the sample space i , the response variable $y_i \in Y \subseteq R^{d_Y}$, the covariate $x_i \in X \subseteq R^{d_X}$, the instrumental variable $z_i \in Z \subseteq R^{d_Z}$, x_i and z_i are strictly exogenous and correlated, $d_X \leq d_Z$, the sequence $\{y_i, z_i\}$ is strictly stationary ;
2. The parameter set is a compact set, let $\theta(\tau) \equiv (\rho(\tau), \beta(\tau)')'$, $\theta_0 \in \Theta \subseteq R^{d_\beta+2}$, $d_\beta \leq d_Z$, given $\tau \in (0,1)$, the parameter $\theta_0(\tau)$ uniquely satisfies the moment condition (9) ;
3. The absolute value of the spatial weight matrix W_n eigenvalue does not exceed 1, the parameter of the spatial lag item $\rho(\tau)$ is in the unit circle, and the matrix $I_n - \rho(\tau)W_n$ non-singular;
4. The matrix $E(z_i * z_i^T)$ is positive definite bounded;
5. Given $u \leq -1$, the bounded function $\widetilde{I}(\cdot)$ satisfies $\widetilde{I}(u) = 0$; $\forall u \geq 1$, the function value is 1 ; and $\forall u \in (-1,1)$, $\widetilde{I}(u) \in [-1,2]$, and satisfy $1 - \int_{-1}^{1} \widetilde{I}^2(u)du > 0$. $\widetilde{I}'(\cdot)$ is of order $r \geq 2$ Bounded symmetric function, satisfying $\int_{-1}^{1} \widetilde{I}' du = 1, \int_{-1}^{1} u^k \widetilde{I}' du = 0$, where

$k=1,2,...,r-1$; and satisfy $\int_{-1}^{1}|u^r\widetilde{I}'|du<\infty, \int_{-1}^{1}|u^{r+1}\widetilde{I}'|du<\infty$, $\int_{-1}^{1}u^r\widetilde{I}'du\neq 0$;

6. The bandwidth h_n satisfies $h_n=o(n^{-1/(2r)})$;
7. Density function $f_{\Lambda|z}(\cdot \mid z)$ exists, bounded and differentiable to order $r,\ r\geq 2$;
8. There is a function $C(z)$, so that $\left|f^{(s)}_{\Lambda|z}(\cdot\mid z)\right|\leq C(z)$, $s=(0,2,...,r)$, for $z\in Z$, $E\left[C(z)\|z\|^2\right]<\infty$;
9. Given an observation z_i, for any fixed parameter $\theta\in\Theta$, Λ_i is continuous at 0;
10. For any given $\tau\in(0,1)$

$$\sup\left\|\hat{M}_n(\theta(\tau),\tau)-E\left[\hat{M}_n(\theta(\tau),\tau)\right]\right\|=o_p(1) \tag{22}$$

11. Let $z_i^{(k)}$ represent the k-th dimension element of z_i, and $\theta^{(j)}$ represent the j-th dimension element of θ , G_{kj} represents the kth row and jth column element of the matrix G:

$$-\frac{1}{nh_n}\sum_{i=1}^{n}\widetilde{I}\left[\frac{-\Lambda_i(Y,X,\tilde{\theta}_k(\tau))}{h_n}\right]z_i^{(k)}\left.\frac{\partial\left[\Lambda_i\left(Y,X,\theta(\tau)\right)\right]}{\partial\theta^{(j)}}\right|_{\theta=\tilde{\theta}_k(\tau)}\xrightarrow{P}G_{kj} \tag{23}$$

Among them, $\forall k,j\in\{1,2,\ldots,d_{\beta+2}\}\,,\theta_0(\tau)\leq\widetilde{\theta}_k(\tau)\leq\hat{\theta}_{\mathrm{MM}}(\tau)$, where $\theta_0(\tau)$ satisfies the moment condition equ::8 , $\hat{\theta}_{\mathrm{MM}}$ satisfies $\hat{M}_n(\hat{\theta}_{\mathrm{MM}},\ tau)=0$;

12. $\Sigma_\tau\equiv\lim_{n\to\infty}(1/n)\ Var\left(g_{ni}(\theta_0(\tau),\tau)\right)$ exists and not singular;
13. $\hat{M}_n(\theta_0(\tau),\tau)$ satisfies the Central Limit Theorem (CLT) , namely

$$\sqrt{n}\left\{\hat{M}_n(\theta_0(\tau),\tau)-E\left[\hat{M}_n(\theta_0(\tau),\tau)\right]\right\}\xrightarrow{d}N(0,\Sigma_\tau) \tag{24}$$

14. The sample weight matrix converges to the population according to probability, $\hat{\Omega}\xrightarrow{P}\Omega>0,\ \|\Omega\|=O(1)$.

Among them, Assumption 1 is the assumption of sample data, together with Assumption 2 it provides conditions for high-level assumptions 10 and 13. Assumption 2 is the global recognition condition, which is the same as in

Hansen (1982)[19] Theorem 2.1 is consistent with the third condition. Considering the stationarity condition, Assumption 3 is the general version of the spatial weight matrix, which is consistent with many spatial econometric literatures. This paper does not consider the optimal selection problem for the time being, and the correlation analysis can refer to Liang (1986)[20]. Compared to the full parameter space $\rho \in P \subseteq R^{d_P}$, assuming 3 only needs ρ_0 to satisfy the uniformly bounded condition , which is consistent with the third assumption in Lee (2001)[21]. Assumption 4 restricts the second-order moments of instrumental variables to satisfy the conditions of the governing convergence theorem, compared to other Quantile regression estimation methods such as 2SLS are weak, which is reflected in Kelejian (1998)[22] and Lee (2003)[23]. Assumption 5 gives conditions for the smooth process of high-order kernel functions, and the condition $1 - \int_{-1}^{1} \widetilde{I}^2(u)du > 0$ ensures the smooth moment function The asymptotic variance of the estimator decreases, consistent with the first and second conditions of the Assumption 4 in Kaplan and Sun (2017)[14]. In citing the law of repeated expectations (LIE), Assumption 7 and 8 ensure the feasibility of calculating the deviation and asymptotic variance of the smooth moment function by using the method of series expansion, and proving the asymptotic normality of the parameters. In Assumption 6, the bandwidth The selection requirement of , in fact, establishes a sufficient condition for the estimator to satisfy asymptotic normality, and this is also the reason why the subscript n is used to represent the bandwidth h_n, which is weaker than Kaplan and Sun (Assumption 5 in 2017)[14] is consistent with assumption 6 in De Castro *et al.* (2019)[15]. To satisfy Vaart Theorem 5.9 in (2000)[24] has certain identification conditions, with assumption 9. Assumption 8 satisfies the stochastic convergence of Theorem 5.9 in Vaart (2000)[24] condition.

Theorem 1. *When assumptions 1-10 and 14 are satisfied, the smoothed generalized moment estimators are consistent.*

$$\hat{\theta}_{\mathrm{SGMM}}(\tau) \xrightarrow{P} \theta_0(\tau) \tag{25}$$

For the detailed proof process, see Appendix A.2.

Theorem 2. *Under the condition that assumptions 1-9 and assumptions 11 to 14 are satisfied, the smoothed generalized moment estimator has asymptotic normality.*

$$\sqrt{n}(\hat{\theta}_{\mathrm{SGMM}}(\tau) - \theta_0(\tau)) \xrightarrow{d} N\left(0, V\right) \tag{26}$$

where $V \equiv (G^\top \Omega\ G)^{-1} G^\top \Omega \Sigma_\tau \Omega\ G\ (G^\top \Omega\ G)^{-1}$.

Theorem 3. *Under the condition that assumptions 1-9 and assumption 14 are satisfied, the smooth generalized moment estimator has asymptotic validity.*

Assumption 14 provides a consistent estimator for the asymptotic variance-covariance matrix. Given that the weight matrix $\Omega \equiv \Sigma_\tau^{-1}$ is asymptotically efficient, the asymptotic covariance matrix, characterized by the difference between $(G^\top \Sigma_\tau^{-1}\ G)^{-1}$ and the covariance matrix estimator of the smoothed generalized method of moments (SGMM), is semi-negative definite. For the detailed proof process, please refer to Appendix A.2 and A.3.

5. Simulation Study

This section focuses on numerically simulating the Smoothed Generalized Moment Method for the Spatial Quantile Regression model, with the aim of assessing its small-sample performance. The comparison involves evaluating both the sample performance and computational speed of this method (SGMM) against the instrumental variable method(IVQR) proposed by Chernozhukov and Hansen (2008) [8]. The assessment criteria include bias (Bias) and root mean square error (RMSE) to provide a comprehensive evaluation of the results $\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}_i - \theta)^2}$

In the formula, m represents the number of simulations, $\hat{\theta}_i$ represents the estimated value of the parameter in the i simulation, and θ represents the known observed value of the parameter. The indicators are all used In order to evaluate the estimation accuracy and robustness of the method, the smaller the index value, the closer the estimated value is to the real value, and the higher the estimation accuracy.

In the formula, m denotes the number of simulations, $\hat{\theta}_i$ represents the estimated value of the parameter in the i-th simulation, and θ represents the known observed value of the parameter. Various indicators are employed to assess the estimation accuracy and robustness of the method. A lower index value indicates closer proximity of the estimated value to the real value, signifying higher estimation accuracy. This comprehensive evaluation provides valuable insights into the performance and reliability of the estimation method.

5.1. The disturbance term is independent and identically distributed

The data generation process (DGP) is based on the following SAR model

$$y_i = \rho \sum_{j=1}^{n} w_{ij} y_j + \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i, \;\; i \neq j \in \{1, 2, ..., n\} \tag{27}$$

Specify the true value of the coefficient as $\rho = 0.7$, $\alpha = 0.1$, $\beta_1 = 0.3$, $\beta_2 = 0.5$, and $x = (x_{1i}, x_{2i})$ is a two-dimensional random variable, which obeys the standard normal distribution and the chi-square distribution with a degree of freedom of 5 and fat-tailed $t(3)$ distribution, respectively. Under the condition of independent and identical distribution of disturbance items, given the quantile level τ , the real value of the parameter can be expressed as $\theta(\tau) = \theta + \Lambda_{\varepsilon_i}^{-1}(\tau)$, where $\Lambda_{\varepsilon_i}^{-1}(\tau)$ is the τ fraction of ε_i. In view of the weight matrices that appear in most empirical literature, including adjacency weight matrix, inverse distance weight matrix, economic weight matrix, and nested matrix, this paper uses the *Rook* adjacency rule to set up the first-order adjacency matrix

$$w_{ij} = \begin{cases} 1, & \text{Spatial unit } i \text{ has adjacent edges to } j; \\ 0, & \text{Spatial units } i \text{ and } j \text{ have no adjacent edges.} \end{cases}$$

Throught the simulation process, various sample sizes $N = \{20, 50, 100, 500\}$ and quantile levels $\tau = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ are considered. A total of 500 simulations are conducted, taking into accout two scenarios where the disturbance term follows a standard normal distribution and a fat-tailed $t(3)$ distribution. Tables 1 and 2 present the deviation and root mean square of the corresponding parameters, providing a comprehensive view of the performance of the error at different quantile levels.

- The SGMM method has a good small sample performance for the spatial quantile regression model;
- The simulation results verified the good large-sample properties of the method;
- In comparison with other quantile levels, the accuracy of the estimated parameters is optimal when $\tau = 0.5$. This observation suggests that the method exhibits promising application potential, particularly in the context of median parameter estimation.

Table 1: The disturbance term is independent and identically distributed in the standard normal distribution

2*	2*	Bias				RMSE			
τ	N	ρ	α	β_1	β_2	ρ	α	β_1	β_2
4*0.1	20	0.0734	-0.5390	0.0256	0.0063	0.4731	4.2775	0.4808	0.1582
	50	0.0201	-0.1338	0.0063	0.0032	0.2037	1.7961	0.2454	0.0923
	100	0.0182	-0.1511	-0.0013	0.0014	0.1096	0.9727	0.1749	0.0601
	500	-0.0010	0.0121	0.0030	-0.0001	0.0457	0.3808	0.0832	0.0250
4*0.3	20	0.0544	-0.3428	0.0030	-0.0153	0.3925	3.3218	0.3247	0.1195
	50	0.0181	-0.1300	-0.0045	-0.0037	0.1458	1.2431	0.2000	0.0659
	100	0.0034	-0.0265	-0.0004	-0.0003	0.0959	0.8309	0.1352	0.0454
	500	0.0021	-0.0186	0.0041	-0.0003	0.0356	0.3018	0.0571	0.0192
4*0.5	20	0.0437	-0.3729	0.0192	0.0012	0.3299	2.9375	0.3129	0.1071
	50	0.0325	-0.2189	-0.0170	-0.0098	0.1323	1.1458	0.1868	0.0627
	100	0.0057	-0.0343	0.0035	-0.0047	0.0863	0.7516	0.1291	0.0436
	500	-0.0008	0.0056	0.0026	0.0006	0.0339	0.2829	0.0563	0.0179

Corresponding to different sample sizes $N = \{20, 50, 100, 500\}$, quantile level $\tau = \{0.1, 0.3, 0.5\}$, compare instrumental variable method (IVQR) and smooth generalized moments The root mean square error of each estimated parameter under the method (SGMM) is shown in the figure below: vertically, compared with the low quantile level, the root mean square error of each parameter at the middle and high quantile level is smaller, when $\tau When$ is close to 0.5, the root mean square error (RMSE) of the smoothed generalized moment method shows a downward trend, and without exception is smaller than the corresponding value under the instrumental variable method; horizontally, as the sample size increases, The instrumental variable method (IVQR) and the smoothed generalized method of moments (SGMM) have increased the accuracy of estimating parameters, which is in line with expectations. In addition, the relative advantage of the latter method has weakened, but still has a comparative advantage; overall, both The performance of small samples is still at the low, middle and high quantile level. The smoothing generalized moment parameter estimation method is always more accurate and more robust than the instrumental variable method, and this advantage is mainly reflected when N is small. It is worth mentioning that the latter This paper will also verify the small-sample computing advantages of the Smoothed Generalized Method of Moments (SGMM), in order to highlight the good application prospects of this method in small samples.

For varying sample sizes $N = \{20, 50, 100, 500\}$ and quantile levels $\tau = \{0.1, 0.3, 0.5\}$, the root mean square error (RMSE) of each estimated parameter is compared between the Instrumental Variable method (IVQR) and the Smoothed Generalized Moments method (SGMM). The figure below illustrates this comparison. Vertically, in comparison to lower quantile levels, the RMSE of each parameter at middle and higher quantile levels is smaller. Particularly, when τ is close to 0.5, the RMSE of the SGMM exhibits a decreasing trend, consistently surpassing the corresponding values under the Instrumental Variable method. Horizontally, as the sample size increases, both the Instrumental Variable method (IVQR) and the Smoothed Generalized Moments method (SGMM) demonstrate improved accuracy in parameter estimation, aligning with expectations. While the relative advantage of the SGMM diminishes with increasing sample size, it still retains a comparative edge. Overall, across low, middle, and high quantile levels, the performance in small samples consistently favors the Smoothed Generalized Moments method. It proves to be more accurate and robust than the Instrumental Variable method, with this advantage prominently pronounced when N is small. Importantly, this paper will further verify the small-sample computational advantages of the Smoothed Generalized Moments method, underscoring its promising application prospects in scenarios with limited data.

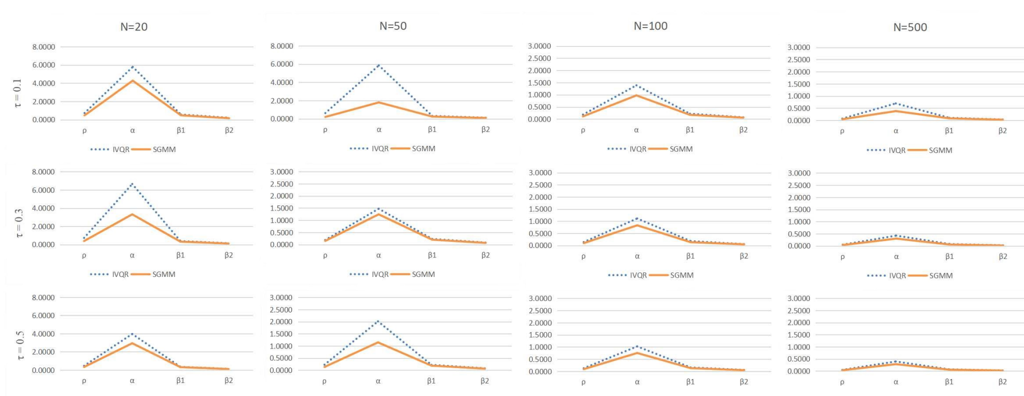

Figure 1: Comparison of the root mean square errors of the two methods under different sample sizes and quantile levels

When the disturbance term follows an independent and identically dis-

Table 2: The disturbance term is independent and identically distributed in $t(3)$ distribution

2*	2*	Bias				RMSE			
τ	N	ρ	α	β_1	β_2	ρ	α	β_1	β_2
4*0.1	20	0.1727	-1.6035	-0.0373	0.0177	0.6869	6.1986	0.7482	0.2545
	50	0.0611	-0.6547	0.0076	-0.0004	0.3479	3.1515	0.4100	0.1471
	100	-0.0010	-0.0728	-0.0193	-0.0072	0.2615	1.8610	0.3091	0.0955
	500	0.0012	-0.0253	-0.0033	-0.0015	0.0743	0.6046	0.1351	0.0405
4*0.3	20	0.0780	-0.6360	0.0016	-0.0111	0.4450	3.7271	0.4410	0.1522
	50	0.0402	-0.3314	0.0039	-0.0071	0.1950	1.6746	0.2419	0.0788
	100	0.0086	-0.0623	-0.0053	-0.0053	0.1078	0.9388	0.1558	0.0534
	500	0.0026	-0.0287	0.0005	0.0001	0.0408	0.3436	0.0731	0.0222
4*0.5	20	0.0568	-0.4974	0.0143	-0.0037	0.3692	3.4926	0.4135	0.1367
	50	0.0340	-0.2722	0.0064	-0.0044	0.1565	1.4346	0.2265	0.0706
	100	0.0098	-0.0608	-0.0131	-0.0048	0.0885	0.7656	0.1435	0.0487
	500	0.0019	-0.0065	-0.0067	-0.0022	0.0364	0.3063	0.0646	0.0201
4*0.7	20	0.0757	-0.6163	-0.0191	-0.0011	0.3823	3.4413	0.4356	0.1463
	50	0.0535	-0.3751	0.0026	-0.0105	0.2071	1.8322	0.2601	0.0782
	100	0.0115	-0.0660	0.0036	-0.0006	0.1405	1.3483	0.1720	0.0516
	500	0.0003	0.0132	-0.0018	-0.0025	0.0407	0.3607	0.0690	0.0220
4*0.9	20	0.1111	-0.8406	0.0529	-0.0222	0.7820	6.9830	0.8387	0.2679
	50	0.0409	-0.2411	-0.0170	-0.0080	0.3284	2.9756	0.4022	0.1423
	100	0.0154	0.0060	0.0092	-0.0117	0.2209	1.9969	0.3064	0.0961
	500	0.0055	-0.0369	-0.0009	0.0001	0.0954	0.9125	0.1376	0.0442

Table 3: Comparison of the root mean square errors of the two methods under different sample sizes and quantile levels

2*τ	2*N	ρ		α		β_1		β_2	
		IVQR	SGMM	IVQR	SGMM	IVQR	SGMM	IVQR	SGMM
4*0.1	20	1.3799	0.6869	10.2095	6.1986	1.0525	0.7482	0.3394	0.2545
	50	0.5997	0.3479	4.1893	3.1515	0.5614	0.4100	0.1681	0.1471
	100	0.3385	0.2615	2.2888	1.8610	0.3638	0.3091	0.1263	0.0955
	500	0.1080	0.0743	0.8855	0.6046	0.1727	0.1351	0.0516	0.0405
4*0.3	20	0.8240	0.4450	7.7205	3.7271	0.6168	0.4410	0.2340	0.1522
	50	0.3429	0.1950	2.6073	1.6746	0.3135	0.2419	0.1104	0.0788
	100	0.2091	0.1078	1.6037	0.9388	0.2222	0.1558	0.0706	0.0534
	500	0.0629	0.0408	0.5104	0.3436	0.1020	0.0731	0.0324	0.0222
4*0.5	20	0.6322	0.3692	5.7060	3.4926	0.5171	0.4135	0.1756	0.1367
	50	0.4194	0.1565	3.9825	1.4346	0.2903	0.2265	0.1024	0.0706
	100	0.1487	0.0885	1.2647	0.7656	0.2071	0.1435	0.0665	0.0487
	500	0.0579	0.0364	0.4898	0.3063	0.0887	0.0646	0.0294	0.0201
4*0.7	20	1.2862	0.3823	10.2983	3.4413	0.5900	0.4356	0.2345	0.1463
	50	0.5777	0.2071	5.4460	1.8322	0.3222	0.2601	0.1156	0.0782
	100	0.1646	0.1405	1.5894	1.3483	0.2187	0.1720	0.0695	0.0516
	500	0.0667	0.0407	0.6572	0.3607	0.0951	0.0690	0.0320	0.0220
4*0.9	20	1.8010	0.7820	16.5050	6.9830	1.0400	0.8387	0.3330	0.2679
	50	0.6685	0.3284	7.5984	2.9756	0.4863	0.4022	0.2415	0.1423
	100	0.5861	0.2209	6.4811	1.9969	0.4220	0.3064	0.1519	0.0961
	500	0.1121	0.0954	1.4018	0.9125	0.1612	0.1376	0.0552	0.0442

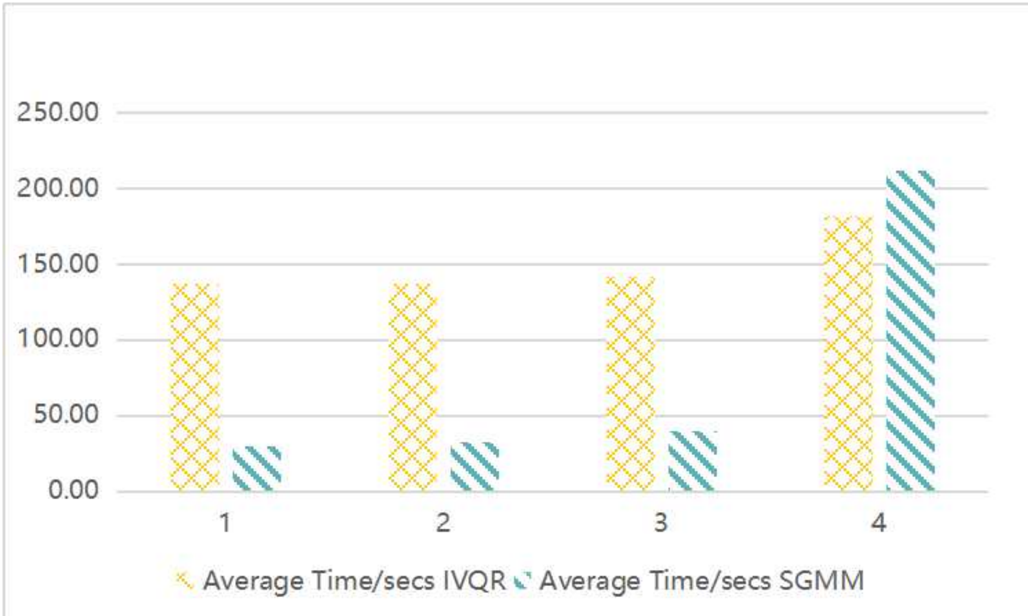

Figure 2: Computational speed comparison

tributed $t(3)$ distribution, the comparison of the two parameter estimation methods is presented in the table above, yielding results similar to those depicted in Figure 6. Comparing the Instrumental Variable method (IVQR) with the computational speed of the Smoothed Generalized Moments method (SGMM), insights can be gleaned from Figure 2. In the context of small samples, the computational advantages of the Smoothed Generalized Moments method are notably apparent. However, as the sample size increases, the computational speed of this method tends to decrease relative to the Instrumental Variable Quantile Regression method. When $n = 500$, the difference between the two is marginal. The computational speed of the Smoothed Generalized Moments method (SGMM) is significantly influenced by the sample size compared to the Instrumental Variable method.

5.2. Nonhomogeneous variance of the disturbance term

If the disturbance term has heteroscedasticity

$$Var(\varepsilon_i) = \sigma^2(x_{1i}, x_{2i}) + \sigma^2(v_i), v_i \sim N(0,1) \tag{28}$$

Other conditions remain unchanged, let $\sigma(x_{1i}, x_{2i}) = 1+0.1x_{1i}+0.2x_{2i}$, then

$$\begin{aligned}
& Q_\tau(y_i - \rho(\tau)\sum_{j=1}^{n} w_{ij}y_j \mid x_{1i}, x_{2i}) && (29) \\
= \; & \alpha(\tau) + \beta_1(\tau)x_{i1} + \beta_2(\tau)x_{i2} + (1 + 0.1x_{1i} + 0.2x_{2i})\,\Lambda_{\varepsilon_i}^{-1}(\tau) \\
= \; & \left[\alpha(\tau) + \Lambda_{\varepsilon_i}^{-1}(\tau)\right] + \left[\beta_1(\tau) + 0.1\Lambda_{\varepsilon_i}^{-1}(\tau)\right] x_{1i} + \left[\beta_2(\tau) + 0.2\Lambda_{\varepsilon_i}^{-1}(\tau)\right] x_{2i} \\
= \; & \alpha_{\varepsilon_i}(\tau) + \beta_{1\varepsilon_i}(\tau) + \beta_{2\varepsilon_i}(\tau) && (30)
\end{aligned}$$

The true value of each parameter at different quantile levels τ can be derived.

In comparison to homoscedasticity, the precision of parameter estimation is relatively diminished under nonhomogeneous variances. However, it still exhibits similar characteristics to those observed under homoscedasticity, as evident in Tables 5.2 and 5.2.

Table 4: The disturbance term is independent and identically distributed in the standard normal distribution

2*	2*	Bias				RMSE			
τ	N	ρ	α	β_1	β_2	ρ	α	β_1	β_2
4*0.1	20	0.0895	-0.8990	-0.0206	0.0804	0.6749	6.1606	0.8342	0.3296
	50	0.0345	-0.4681	-0.0267	0.0335	0.4387	3.3744	0.4790	0.2046
	100	0.0260	-0.2889	0.0192	0.0118	0.2463	1.9316	0.3463	0.1374
	500	-0.0028	0.0028	0.0105	0.0006	0.0976	0.7802	0.1370	0.0640
4*0.3	20	0.1189	-0.9243	-0.0454	0.0205	0.5436	4.9068	0.6287	0.2629
	50	0.0862	-0.6981	-0.0061	0.0000	0.2426	2.2157	0.3669	0.1572
	100	0.0418	-0.3691	0.0009	0.0052	0.1589	1.3644	0.2453	0.1000
	500	0.0071	-0.0654	0.0006	0.0009	0.0634	0.5143	0.1054	0.0477
4*0.5	20	0.1335	-1.0936	-0.0440	-0.0083	0.5132	4.4611	0.5923	0.2523
	50	0.0673	-0.5045	-0.0401	-0.0062	0.2513	2.1747	0.3417	0.1473
	100	0.0350	-0.2685	0.0032	-0.0003	0.1763	1.4717	0.2475	0.1048
	500	0.0016	-0.0153	0.0008	0.0013	0.0610	0.5004	0.0981	0.0448

Similarly, comparing the performance of the two parameter estimation methods under different quantile levels and sample sizes when the variance of the disturbance term is not homogeneous, the root mean square error has increased compared to the case of the same variance, but the comparison results of the methods Consistent with figure 2, see figure 3 and table 6.

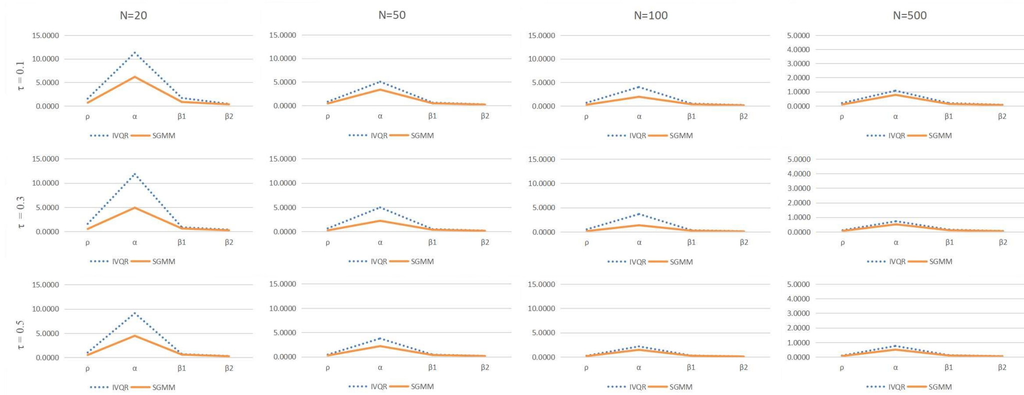

Figure 3: Comparison of the root mean square errors of the two methods under different sample sizes and quantile levels

Similarly, when examining the performance of the two parameter estimation methods under different quantile levels and sample sizes in the presence of nonhomogeneous variance in the disturbance term, the root mean square error experiences an increase compared to the scenario with homogeneous variance. However, the comparative results of the methods remain consistent with those depicted in Figure 2, as illustrated in Figure 3 and Table 5.2.

6. Empirical Research

Referring to Huang Dahu (2021)[29], economic growth is the most direct and effective way to achieve poverty reduction. In the process of economic development, labor is transferred between regions and sectors. The transfer of rural labor force is not only the product of economic development to a certain stage, but also an important force to promote economic growth. Economic growth can effectively alleviate poverty, so the transfer of rural labor can affect poverty. Economic growth has a diffusion effect, that is, the economic growth of a region can drive the economic development of adjacent regions, thereby helping to alleviate rural poverty in adjacent regions. This part aims to describe the heterogeneity of the spatial aggregation effect of rural relative poverty and the spatial heterogeneity of rural labor transfer for poverty reduction by constructing a spatial quantile regression model and using the smoothed generalized moment estimation method to estimate parameters.

Table 5: The disturbance term is independent and identically distributed in $t(3)$ distribution

2*	2*	Bias				RMSE			
τ	N	ρ	α	β_1	β_2	ρ	α	β_1	β_2
4*0.1	20	0.2612	-2.8912	-0.0056	0.0695	1.1017	11.5424	1.7963	0.6457
	50	0.1428	-1.8255	0.0179	0.0738	0.5296	4.8926	0.7851	0.3420
	100	0.0234	-1.0895	-0.0042	0.0355	0.6600	3.7346	0.6159	0.2555
	500	-0.0205	-0.1394	-0.0169	0.0123	0.3093	1.6329	0.2389	0.1105
4*0.3	20	0.1436	-1.4064	0.0137	0.0238	0.5874	5.5913	0.8892	0.3487
	50	0.0942	-0.8915	0.0173	0.0163	0.2623	2.3513	0.4560	0.1847
	100	0.0543	-0.5446	0.0126	0.0060	0.3241	2.7581	0.3059	0.1336
	500	0.0067	-0.0683	-0.0019	-0.0030	0.0881	0.7144	0.1338	0.0552
4*0.5	20	0.1145	-0.8702	0.0303	-0.0007	0.4325	3.8415	0.7608	0.2908
	50	0.0873	-0.6871	0.0086	-0.0135	0.2945	2.4277	0.4273	0.1682
	100	0.0423	-0.3377	0.0145	0.0029	0.2214	2.0097	0.2726	0.1159
	500	0.0077	-0.0672	-0.0020	0.0013	0.0772	0.6747	0.1130	0.0492
4*0.7	20	0.1351	-1.0627	-0.0137	-0.0161	0.5888	5.9051	0.7980	0.3177
	50	0.1001	-0.6611	-0.0042	-0.0248	0.2881	2.3226	0.4787	0.1912
	100	0.0543	-0.3094	-0.0080	-0.0186	0.2353	1.9140	0.3178	0.1346
	500	0.0112	-0.0601	-0.0087	-0.0044	0.0751	0.6679	0.1250	0.0571
4*0.9	20	0.2793	-1.5116	-0.1877	-0.1325	1.0308	9.4241	1.6580	0.6044
	50	0.1566	-0.4168	-0.0691	-0.0884	0.6124	5.4255	0.9131	0.3760
	100	0.1492	-0.6121	-0.0279	-0.0610	0.5227	3.7071	0.5654	0.2733
	500	0.0401	-0.1086	-0.0352	-0.0238	0.1974	1.7530	0.2441	0.1241

Table 6: Comparison of the root mean square errors of the two methods under different sample sizes and quantile levels

2*τ	2*N	ρ		α		β_1		β_2	
		IVQR	SGMM	IVQR	SGMM	IVQR	SGMM	IVQR	SGMM
4*0.1	20	2.9234	1.1017	30.8135	11.5424	2.7991	1.7963	0.9159	0.6457
	50	1.0483	0.5296	5.2753	4.8926	1.0171	0.7851	0.4123	0.3420
	100	0.7710	0.6600	3.9526	3.7346	0.7833	0.6159	0.3003	0.2555
	500	0.5506	0.3093	2.2171	1.6329	0.3698	0.2389	0.1390	0.1105
4*0.3	20	1.0850	0.5874	10.0630	5.5913	1.0340	0.8892	0.4490	0.3487
	50	0.7213	0.2623	6.1143	2.3513	0.6609	0.4560	0.2588	0.1847
	100	0.5436	0.3241	3.6331	2.7581	0.4543	0.3059	0.1791	0.1336
	500	0.1598	0.0881	1.0542	0.7144	0.1849	0.1338	0.0724	0.0552
4*0.5	20	1.3906	0.4325	11.8639	3.8415	0.9683	0.7608	0.3516	0.2908
	50	0.5860	0.2945	5.0936	2.4277	0.5891	0.4273	0.2077	0.1682
	100	0.5025	0.2214	4.3642	2.0097	0.3690	0.2726	0.1478	0.1159
	500	0.1267	0.0772	1.0478	0.6747	0.1687	0.1130	0.0653	0.0492
4*0.7	20	2.1114	0.5888	18.4968	5.9051	1.2764	0.7980	0.3871	0.3177
	50	0.6978	0.2881	7.4463	2.3226	0.6322	0.4787	0.2460	0.1912
	100	0.3680	0.2353	3.9148	1.9140	0.4410	0.3178	0.1716	0.1346
	500	0.1189	0.0751	1.1767	0.6679	0.1890	0.1250	0.0720	0.0571
4*0.9	20	1.7051	1.0308	20.4606	9.4241	2.1522	1.6580	0.6965	0.6044
	50	1.5269	0.6124	14.5664	5.4255	1.1363	0.9131	0.5625	0.3760
	100	0.4450	0.5227	5.8879	3.7071	0.7307	0.5654	0.3173	0.2733
	500	0.1902	0.1974	2.4306	1.7530	0.3167	0.2441	0.1361	0.1241

Table 7: Descriptive statistics for selected indicators

Variable	Mean	SD	Minmum	25%	Median	75%	Maximum
IoP	0.3287	0.0403	0.2780	0.2970	0.3160	0.3445	0.4380
LT	0.5610	0.0517	0.4619	0.5160	0.5601	0.5948	0.6482
ID	2.3023	0.6429	0.2189	2.2227	2.4015	2.5119	3.2696
EG	2.5311	1.0300	0.9626	1.9906	2.3014	2.9537	5.5585
FI	1.0136	0.5668	0.0963	0.5656	1.0844	1.3641	2.1941
RoU	0.6373	0.1106	0.3573	0.5753	0.6264	0.6911	0.8930

6.1. Data source and index selection

To investigate the heterogeneous impact of rural labor force on rural poverty, the core explanatory variable chosen is the degree of rural labor transfer (Labor Transfer or LT). The incidence of rural poverty (IoP) is selected as the explained variable, and relevant control variables are considered to enhance the robustness of the model. The selection rationale for each variable is as follows:

1. Explained Variable: The incidence of rural poverty (IoP) is measured using the rural Engel coefficient of each province.

2. Core Variable: The degree of rural labor force transfer (LT) is represented by the ratio of rural employees not engaged in the primary industry to the total number of rural employees, serving as a proxy variable to quantify the extent of rural labor force transfer.

3. Control Variables: Urban-rural Income Gap (ID): The ratio of per capita disposable income of urban residents to rural residents, employed to measure the income disparity between urban and rural areas. Rural Economic Development Level (EG): This variable is gauged by the per capita total output value of rural agriculture, forestry, animal husbandry, and fishery in each province. Infrastructure (FI): Measured as the ratio of road mileage in each province to the province's area, reflecting road density. Rate of Urbanization (RU): Calculated by the proportion of the urban population to the total population, serving as an indicator of urbanization development.

The data utilized in this study are sourced from various statistical yearbooks, including the "China Statistical Yearbook," "China Rural Statistical Yearbook," and "China Population and Employment Statistical Yearbook." The analysis focuses on the data from 2021 across 31 provincial-level regions in mainland China. A detailed statistical description of each variable

Table 8: Descriptive statistics for selected indicators

	IoP	LT	ID	EG	FI	RoU
count	31	31	31	31	31	31
mean	0.3287	0.5610	2.3023	2.5311	1.0136	0.6373
std	0.0403	0.0517	0.6429	1.0300	0.5668	0.1106
min	0.2780	0.4619	0.2189	0.9626	0.0963	0.3573
25%	0.2970	0.5160	2.2227	1.9906	0.5656	0.5753
50%	0.3160	0.5601	2.4015	2.3014	1.0844	0.6264
75%	0.3445	0.5948	2.5119	2.9537	1.3641	0.6911
max	0.4380	0.6482	3.2696	5.5585	2.1941	0.8930

is presented in Table 6.1.

6.2. Spatial weight matrix selection and model specify

Figure 4 illustrates the estimation outcomes of the spatial autoregressive model employing different k adjacent spatial weight matrices. The shaded region signifies instances where the regression coefficient is non-stationary. To ensure the spatial stability of the autoregressive model, this study opts for 8 neighbors, employing the Queen adjacency criterion to derive the spatial weight matrix for prefecture-level cities. Subsequently, the paper establishes the spatial quantile regression, employing the aforementioned smoothing technique on the objective function to yield consistent parameter estimates. The detailed results are presented in Table 6.1.

$$\begin{aligned} Q_\tau(IoP_i) \quad = \quad & \rho(\tau)\sum_{j\neq i} w_{ij} IoP_j + \alpha(\tau) + \beta_1(\tau)LT_i + \beta_2(\tau)ID_i + \beta_3(\tau)EG_i \\ & + \beta_4(\tau)FI_i + \beta_5(\tau)RoU_i \end{aligned}$$

where w_{ij} represents the element of the spatial weight matrix.

6.3. Model Analysis Results

Based on the above analysis, combined with Figure 5, it is not difficult to draw the following conclusions: (1) Consistent with previous scholars' research conclusions [26,29], the incidence of poverty in rural areas has a significant spatial agglomeration effect, showing The "arch" clustering trend, that is, the spatial effect at the extreme quantile levels of 0.1 and 0.9

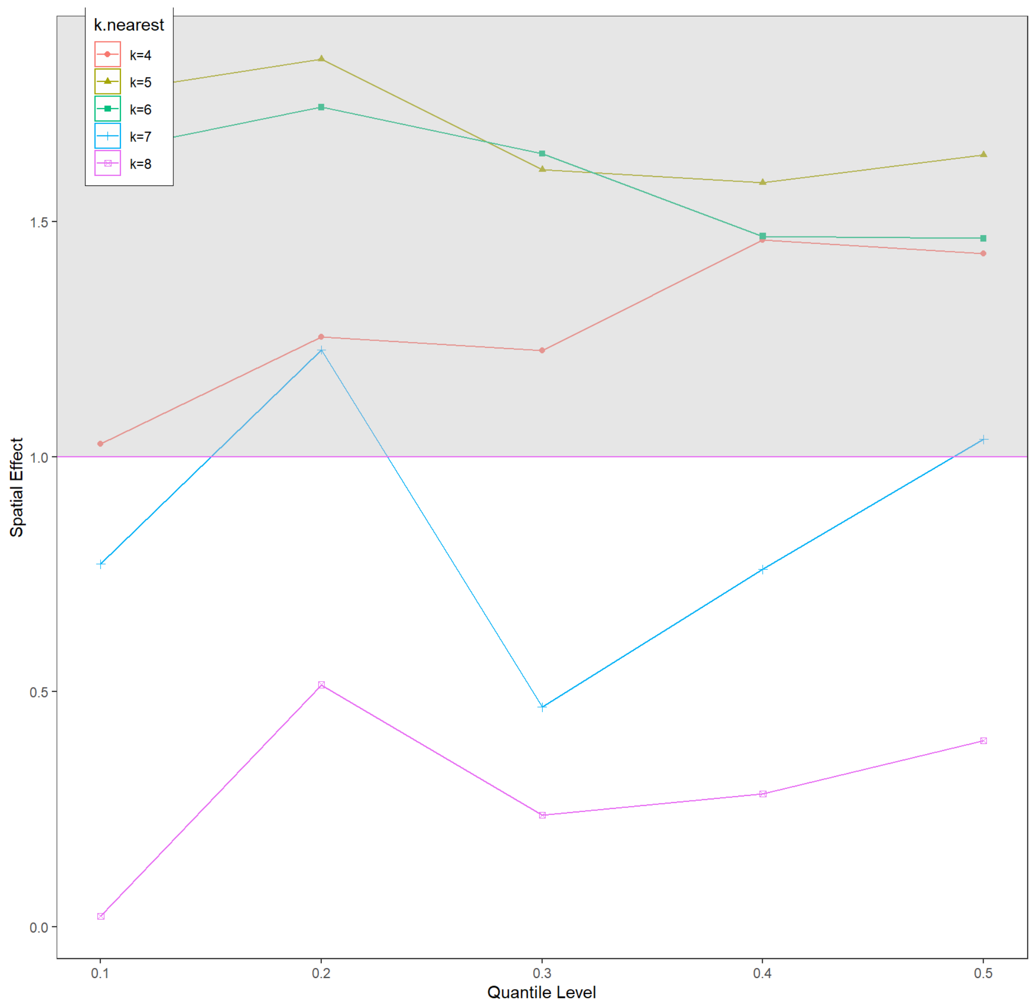

Figure 4: Stationarity of spatial autoregressive models with different k neighborhood connections

Table 9: Empirical Analysis Results

	τ=0.1	τ=0.2	τ=0.4	τ=0.5	τ=0.6	τ=0.8	τ=0.9
$\rho(\tau)$	0.022903	0.51466	0.282745	0.395762	0.295813	0.38804	-0.21643
$\alpha(\tau)$	0.073573	0.009389	0.020803	-0.01723	-0.02627	0.10125	0.31115
$\beta_1(\tau)$	-0.02866	-0.03906	-0.03535	-0.04826	-0.04772	-0.02774	0.014756
$\beta_2(\tau)$	0.0149	-0.00128	-0.00233	-0.00811	-0.00868	0.013	0.049227
$\beta_3(\tau)$	-0.02574	-0.01645	-0.00464	0.005099	0.005741	0.02465	0.008171
$\beta_4(\tau)$	-0.0006	-0.06902	-0.13537	-0.18154	-0.1748	-0.29276	-0.16945
$\beta_5(\tau)$	0.310937	0.288293	0.394913	0.44831	0.481274	0.38507	0.280878

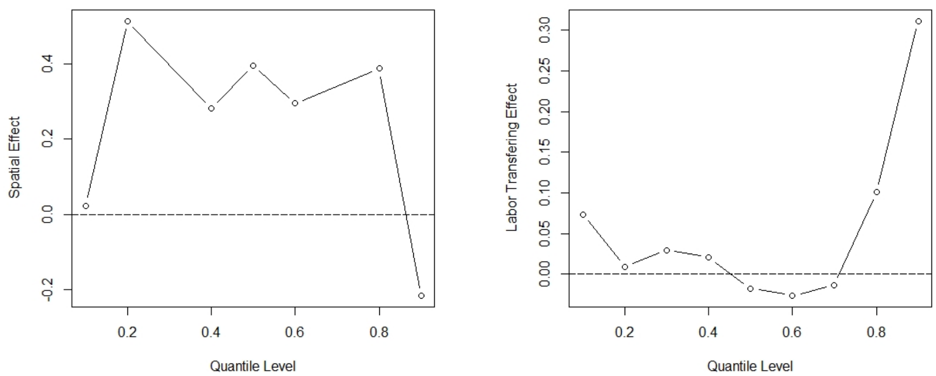

Figure 5: Heterogeneity at different quantile levels

is smaller than the rest of the quantile points, and its spatial effect turns from positive to negative at the high quantile level; Rural areas with a relatively high rate of poverty often present a "dispersed" layout, which also proves the remarkable achievements of China's poverty alleviation from another perspective. (2) Consistent with the economic mechanism [26], Rural labor transfers have significant spatial spillover effects at higher quantile levels, that is, in provincial rural areas with high poverty incidence, higher rural labor transfers aggravate poverty incidence; but this conclusion may not be applicable to other quantile levels , for example, at the level of $\tau = 0.6$, the transfer of rural labor force will be conducive to rural poverty reduction.

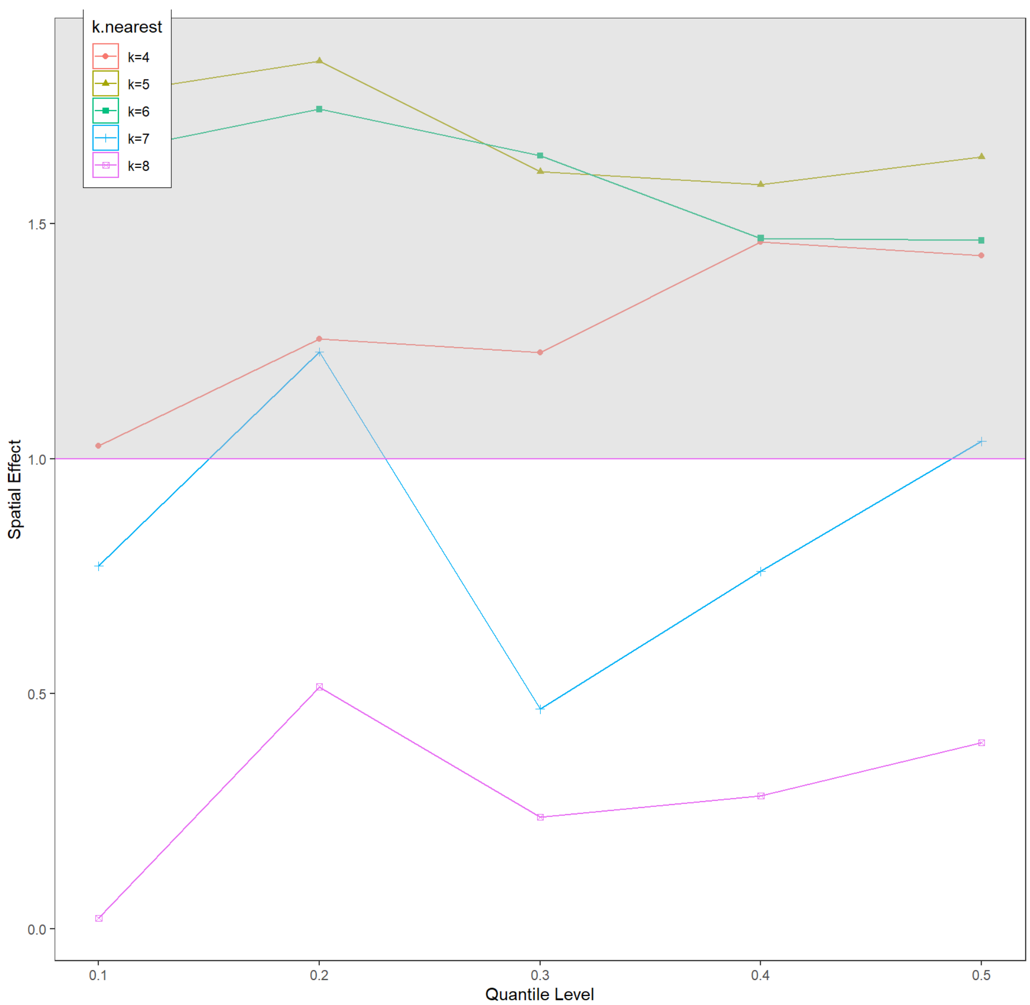

Figure 6: Stationarity of spatial autoregressive models with different k neighborhood connections

7. Conclusion

This paper employs Smoothed Generalized Moment Estimation to investigate the Spatial Quantile Regression model, effectively addressing challenges related to endogeneity and the non-differentiability of the objective function. The optimization algorithm, guided by the first-order condition, yields the Smoothed Generalized Moment Estimator. In comparison to both the unsmoothed counterpart and the Smooth Quantile Regression Loss function (Horowitz, 1998) [16], the estimator exhibits computational and testing advantages. The Smoothed Generalized Moment Estimator is established as consistent, asymptotically normal, and asymptotically valid. Numerical simulations showcase its superior small-sample performance and computational speed. In empirical studies, parameter estimation using Smoothed Generalized Moments in the Spatial Quantile Regression model effectively captures the heterogeneity of spatial aggregation effects on rural relative poverty and the spatial heterogeneity of rural labor transfer for poverty reduction.

Reference

[1] Ripley B D, Cliff A D, Ord J K, et al. Spatial processes: models and applications [J]. *Journal of the American Statistical Association*, 1984, 147(3): 515-515.

[2] Anselin L, Bera A.K. Spatial dependence in linear regression models with an introduction to spatial econometrics [M]. *Statistics: Textbooks and Monographs*,1998, 155: 237-289.

[3] Frissen A, Lieverse R, Drukker M, et al. Evidence that childhood urban environment is associated with blunted stress reactivity across groups of patientswith psychosis, relatives of patients and controls [J]. *Social Psychiatry and Psychiatric Epidemiology*, 2014, 49(10): 1579-1587.

[4] Qu X, Lee L F. Estimating a spatial autoregressive model with an endogenous spatial weight matrix [J]. *Journal of Econometrics*, 2015, 184(2): 209-232.

[5] Jin F, Lee L F. Outer-product-of-gradients tests for spatial autoregressive models [J]. *Regional ence and urban economics*, 2018, 72(SEP.):35-57.

[6] Koenker R. Quantile regression [J]. *Journal of Economic Perspectives*, 2005, 15(4):143-156.

[7] Kim, TaeHwan, Muller, Christophe. Two-stage quantile regression when the first stage is based on quantile regression [M]. *The Econometrics Journal*, 2004.

[8] Chernozhukov V, Hansen C. Instrumental variable quantile regression: A robust inference approach [J]. *Journal of Econometrics*, 2008, 142(1): 379-398.

[9] Su L, Yang Z. Instrumental variable quantile estimation of spatial autoregressive models [J]. *Working Papers*, 2007.

[10] McMillen D P. Quantile regression for spatial data [M]. *Springer Science & Business Media*, 2012.

[11] Li K, Fang L. Instrumental Variables Estimation and Parametric Tests of Spatial Lag Quantile Regression Model[J]. *Statistical Research*, 2018, 35(10): 103-115.

[12] Zietz J, Zietz E N, Sirmans G S. Determinants of house prices: A quantile regression approach [J]. *The Journal of real estate finance and economics*, 2008, 37(4): 317-333.

[13] Kostov P. Spatial dependence in agricultural land prices: Does it exist? [J]. *Agricultural Economics*, 2009, 40(3): 347-353.

[14] Kaplan D M, Sun Y. Smoothed estimating equations for instrumental variables quantile regression [J]. *Econometric Theory*, 2017, 33(01): 105-157.

[15] De Castro L I, Galvao A F, Kaplan D M, et al. Smoothed GMM for quantile models [J]. *Journal of Econometrics*, 2019, 213(1): 121-144.

[16] Horowitz J L. Bootstrap methods for median regression models [J]. *Econometrica*, 1998: 1327-1351.

[17] Liu X, Lee L, Bollinger C R. An efficient GMM estimator of spatial autoregressive models [J]. *Journal of Econometrics*, 2010, 159(2): 303-319.

[18] Chernozhukov V, Hansen C. An IV model of quantile treatment effects [J]. *Econometrica*, 2005, 73(1): 245-261.

[19] Hansen L P. Large sample properties of generalized method of moments estimators [J]. *Econometrica: Journal of the Econometric Society*, 1982: 1029-1054.

[20] Liang K Y, Zeger S L. Longitudinal data analysis using generalized linear models [J]. *Biometrika*, 1986, 73(1): 13-22.

[21] Lee L. Generalized method of moments estimation of spatial autoregressive processes [J]. *Unpublished manuscript*, 2001.

[22] Kelejian H H, Prucha I R. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances [J]. *The Journal of Real Estate Finance and Economics*, 1998, 17(1):99-121.

[23] Lee L F. Best Spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances [J]. *Econometric Reviews*, 2003, 22(4):307-335.

[24] Van der Vaart A W. Asymptotic statistics [M]. *Cambridge university press*, 2000.

[25] White H. Estimation, inference and specification analysis [M]. *Cambridge university press*, 1996.

[26] Zhang G, Wang Q, Zhang R. Research on the poverty reduction effect of agricultural labor transfer in China[J]. *Chinese Journal of Population Science*, 2018 (04): 18-29+126.

[27] Cai F, Wang D. Sustainability of China's economic growth and labor contribution[J]. *Economic Research Journal*, 1999 (10): 62-68.

[28] Zhang B, Yang Z. Rural poverty reduction and its spatial spillover effects of Chinese urbanization: Based on the analysis of spatial econometric model with provincial panel data[J]. *Geographical Research*,2020, 39 (07): 1592-1608.

[29] Huang D, Ding S, Tan C. Poverty Reduction And Spatial Spillover Effect Of Rural Labor Transfer[J]. *Chinese Journal of Agricultural Resources and Regional Planning*, 2022,43(04):227-236.

[30] Hong M, Pan D, Wu Z, He Y. THE EFFICIENCY OF RURAL POVERTY REDUCTION AND ITS SPATIAL DISPARITIES ———BASED ON THE DEA ESDA MODEL[J]. *Chinese Journal of Agricultural Resources and Regional Planning*,

[31] Shan D, Zhang C, Wang Y. Rural-urban Poverty MigrationUrbanization Mode and Heterogeneous Poverty Reduction Effect[J]. *China Population,Resources and Environment*, 2015,25 (09): 81-92.

[32] He C, Cui W. The mechanism of rural labor transfer to reduce poverty[J]. *Urban Problems*, 2018,(03): 27-33.

[33] Kato K. Asymptotic normality of Powell's kernel estimator [J]. *Annals of the Institute of Statistical Mathematics*, 2012, 64(2): 255-273.

Appendix .1. A.1 Method Advantages

$$\Lambda_i(Y,X,\theta(\tau)) \equiv y_i - \left[\rho(\tau)\sum_{j=1}^{n} w_{ij}y_j + x_i'\beta(\tau)\right] \tag{.1}$$

Note that the marginal distribution density function of Λ_i is $f_\Lambda(\cdot)$, and the conditional distribution density function is $f_{\Lambda|z}(\cdot \mid z_i)$. Given $z_i = z$, assume its value range is $[\Lambda_L(z), \Lambda_H(z)]$, and satisfy $\Lambda_L(z) \leq -h_n \leq h_n \leq \Lambda_H(z)$. has

$$\begin{aligned} & E\left[g_{ni}(\theta_0(\tau),\tau)\right] \\ = \; & E\left\{E\left[\widetilde{I}(-\Lambda_i(Y,X,\theta_0(\tau))/h_n) - \tau \mid z_i\right] z_i\right\} \\ = \; & E\left\{z_i \int_{\Lambda_L(z_i)}^{\Lambda_H(z_i)} \left[\widetilde{I}(-x/h_n) - \tau\right] dF_{\Lambda|z}(x \mid z_i)\right\} \end{aligned} \tag{.2}$$

According to Kaplan and Sun (2017)[14], based on the assumption of the value range of Λ_i, combined with the Assumption 7, using the integration by parts method and expanding it into a series, we can get have to

$$\begin{aligned} & E\left[g_{ni}(\theta_0(\tau),\tau)\right] \\ = \; & E\left\{z_i\left[-\tau + h_n^{-1}\int_{-1}^{1} F_{\Lambda|z(-h_n\upsilon|z_i)}\widetilde{I}'(\upsilon)d\upsilon\right]\right\} \\ = \; & E\left\{z_i\left[-\tau + \sum_{k=0}^{r} F_{\Lambda|z}^{(k)}(0 \mid z_i)\frac{(-h_n)^k}{k!}\int_{-1}^{1}\upsilon^k\widetilde{I}'(\upsilon)d\upsilon\right]\right\} \\ + \; & O(h_n^{r+1})E\left[z_i\int_{-1}^{1} f_{\Lambda|z}^{(r)}(-\widetilde{h}\upsilon \mid z_i)\upsilon^{r+1}\widetilde{I}'(\upsilon)d\upsilon\right] \end{aligned} \tag{.3}$$

Based on Assumption 8 , there exists a bounded function $C(\cdot)$, with

$$E\left[z_i\int_{-1}^{1} f_{\Lambda|z}^{(r)}(-\widetilde{h}\upsilon \mid z_i)\upsilon^{r+1}\widetilde{I}'(\upsilon)d\upsilon\right] \leq E\left[\int_{-1}^{1} C(z) \parallel z \parallel \mid \upsilon^{r+1}\widetilde{I}'(\upsilon) \mid d\upsilon\right] < \infty \tag{.4}$$

Based on Assumption 5 , the formula (.3) can be transformed into

$$\begin{aligned} & E\left[g_{ni}(\theta_0(\tau),\tau)\right] \\ = \; & E\left\{z_i\left[-\tau + F_{\Lambda|z}(0 \mid z_i) + f_{\Lambda|z}^{(r-1)}(0 \mid z_i)\frac{(-h_n)^r}{r!}\int_{-1}^{1}\upsilon^r\widetilde{I}'(\upsilon)d\upsilon\right]\right\} + O(h_n^{r+1}) \end{aligned}$$

$$= E\left\{z_i\left[-\tau + E(I\{\Lambda_i \le 0\} \mid z_i)\right]\right\} + \frac{(-h_n)^r}{r!}\left[\int_{-1}^{1} \upsilon^r \widetilde{I}'(\upsilon)d\upsilon\right] E\left[z_i f_{\Lambda|z}^{(r-1)}(0 \mid z_i)\right] + O(h_n^{r+1}) \tag{.5}$$

Given $E\left[z_i(I\{\Lambda_i \le 0\} - \tau) \mid z_i\right] = z_i\left[-\tau + E(I\{\Lambda_i \le 0\} \mid z_i)\right]$, and the value of the former is zero, combined with the assumption 8, the above formula can be transformed into

$$E\left[g_{ni}(\theta_0(\tau), \tau)\right] = \frac{h_n^r}{r!}\left[\int_{-1}^{1} \upsilon^r \widetilde{I}'(\upsilon)d\upsilon\right] E\left[z_i f_{\Lambda|z}^{(r-1)}(0 \mid z_i)\right] + O(h_n^{r+1}) \tag{.6}$$

From this, we get the display expression of the estimated deviation of the smooth moment function, and it is not difficult to see that it is actually a high-order infinitesimal quantity of h^r.

Then calculate the asymptotic variance of the smooth moment function

$$\begin{aligned} & E\left[g_{ni}'(\theta_0(\tau), \tau) g_{ni}(\theta_0(\tau), \tau)\right] \\ = \; & E\left\{E\left[\widetilde{I}(-\Lambda_i(Y, X, \theta_0(\tau))/h_n) - \tau \mid z_i\right] z_i' z_i\right\} \\ = \; & E\left\{z_i' z_i \int_{\Lambda_L(z_i)}^{\Lambda_H(z_i)} \left[\widetilde{I}(-x/h_n) - \tau\right]^2 dF_{\Lambda|z}(x \mid z_i)\right\} \end{aligned} \tag{.7}$$

where

$$\begin{aligned} & \int_{\Lambda_L(z_i)}^{\Lambda_H(z_i)} \left[\widetilde{I}(-x/h_n) - \tau\right]^2 dF_{\Lambda|z}(x \mid z_i) \\ = \; & \left[\widetilde{I}(-x/h_n) - \tau\right]^2 F_{\Lambda|z}(x \mid z_i) \,\Big|_{\Lambda_L(z_i)}^{\Lambda_H(z_i)} \\ + \; & \frac{2}{h_n} \int_{\Lambda_L(z_i)}^{\Lambda_H(z_i)} F_{\Lambda|z}(x \mid z_i)\left[\widetilde{I}(-x/h_n) - \tau\right]\left[\widetilde{I}'(-x/h_n)\right] dx \end{aligned} \tag{.8}$$

Based on the Assumption 5 and Assumption of the range of Λ_i, combined with the Assumption 7 , using the method of integration by parts and expanding it into a series, similarly we can get

$$\begin{aligned} & \int_{\Lambda_L(z_i)}^{\Lambda_H(z_i)} \left[\widetilde{I}(-x/h_n) - \tau\right]^2 dF_{\Lambda|z}(x \mid z_i) \\ = \; & \tau^2 + 2\int_{-1}^{1} F_{\Lambda|z}(h_n \upsilon \mid z_i)\left[\widetilde{I}(-\upsilon) - \tau\right]\left[\widetilde{I}'(-\upsilon)\right] d\upsilon \\ = \; & \tau^2 + \tau C + h_n f_{\Lambda|z}(0 \mid z_i) D + h_n^2 E \end{aligned} \tag{.9}$$

where mark

$$
\begin{aligned}
C &\equiv 2\int_{-1}^{1}\left[\widetilde{I}(-\upsilon)-\tau\right]\left[\widetilde{I}'(-\upsilon)\right]d\upsilon \\
&= 2\int_{-1}^{1}\left[\widetilde{I}(\upsilon)-\tau\right]\left[\widetilde{I}'(\upsilon)\right]d\upsilon \\
&= \left[\widetilde{I}^2(\upsilon)-2\tau\widetilde{I}(\upsilon)\right]|_{-1}^{1} \\
&= 1-2\tau \qquad (.10)
\end{aligned}
$$

$$
\begin{aligned}
D &\equiv 2\int_{-1}^{1}\upsilon\left[\widetilde{I}(-\upsilon)-\tau\right]\left[\widetilde{I}'(-\upsilon)\right]d\upsilon \\
&= -2\int_{-1}^{1}\upsilon\left[\widetilde{I}(\upsilon)-\tau\right]\left[\widetilde{I}'(\upsilon)\right]d\upsilon \\
&= -\left[\upsilon\widetilde{I}^2(\upsilon)\,|_{-1}^{1}-\int_{-1}^{1}\widetilde{I}^2(\upsilon)d\upsilon\right] \\
&= -\left[1-\int_{-1}^{1}\widetilde{I}^2(\upsilon)d\upsilon\right] \qquad (.11)
\end{aligned}
$$

$$
\begin{aligned}
|E| &\equiv \left|\int_{-1}^{1}\upsilon^2 f'_{\Lambda|z}(\tilde{h_n}\upsilon \mid z_i)\left[\widetilde{I}(-\upsilon)-\tau\right]\left[\widetilde{I}'(-\upsilon)\right]d\upsilon\right| \\
&\leq \int_{-1}^{1}C(z_i)\left|\upsilon^2\left[\widetilde{I}'(-\upsilon)\right]\right|<\infty \qquad (.12)
\end{aligned}
$$

Therefore, the formula (.7), that is, the display expression of the asymptotic variance is

$$
\begin{aligned}
& E(g'_{ni}(\theta_0(\tau),\tau)g_{ni}(\theta_0(\tau),\tau)) \\
=\; & \tau(1-\tau)E[z'_i z_i]-h_n\Big[1-\int_{-1}^{1}\widetilde{I}^2(u)du\Big]E[f_{\Lambda|z}(0|z_i)z'_i z_i]+O(h_n^2)
\end{aligned}
$$

in the same way

$$
\begin{aligned}
& E(g_{ni}(\theta_0(\tau),\tau)g'_{ni}(\theta_0(\tau),\tau)) \\
=\; & \tau(1-\tau)E[z_i z'_i]-h_n\Big[1-\int_{-1}^{1}\widetilde{I}^2(u)du\Big]E[f_{\Lambda|z}(0|z_i)z_i z'_i]+O(h_n^2) \qquad (.13)
\end{aligned}
$$

Appendix .2. Consistency

First prove the consistency of the initial value of the estimated parameters $\hat{\theta}_{\mathrm{MM}}$ participating in the iteration. The theorem in van der Vaart (2000)[24] On the basis of 5.9, it is easy to prove, $|Z[\widetilde{I}(\cdot - I\{\cdot\})]| \leq |Z|$, based on Assumption 4 , Assumption 10 and control Convergence theorem, there is

$$\begin{aligned}
& \lim_{h_n \to 0} \sup \left\| E\left[\hat{M}_n(\theta(\tau),\tau)\right] - M^u(\theta(\tau),\tau) \right\| \\
\equiv\ & \lim_{h_n \to 0} \sup \left\| E\left[\hat{M}_n(\theta(\tau),\tau)\right] - E\left\{z_i\left[I\left\{\Lambda_i(Y,X,\theta(\tau)) \leq 0\right\} - \tau\right]\right\} \right\| \\
=\ & \lim_{h_n \to 0} \max \left\| E\left\{ z_i \left[\widetilde{I}\left(\frac{-\Lambda_i(Y,X,\theta(\tau))}{h_n} \right) - I\left\{\Lambda_i(Y,X,\theta(\tau)) \leq 0\right\} \right] \right\} \right\| \\
=\ & \lim_{h_n \to 0} \left\| E\left\{ z_i \left[\widetilde{I}\left(\frac{-\Lambda_i(Y,X,\theta^*(\tau))}{h_n} \right) - I\left\{\Lambda_i(Y,X,\theta^*(\tau)) \leq 0\right\} \right] \right\} \right\| \\
=\ & \left\| E\left\{ \lim_{h_n \to 0} z_i \left[\widetilde{I}\left(\frac{-\Lambda_i(Y,X,\theta^*(\tau))}{h_n} \right) - I\left\{\Lambda_i(Y,X,\theta^*(\tau)) \leq 0\right\} \right] \right\} \right\| \\
=\ & 0 \qquad (.14)
\end{aligned}$$

Based on the continuity of the moment function, $\theta^*(\tau)$ represents the value when the above formula reaches the maximum. Therefore, combining the triangle inequality and Assumption 10

$$\begin{aligned}
& \sup \left\| \hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau) \right\| \\
=\ & \sup \left\| \hat{M}_n(\theta(\tau),\tau) - E\left[\hat{M}_n(\theta(\tau),\tau)\right] + E\left[\hat{M}_n(\theta(\tau),\tau)\right] - M^u(\theta(\tau),\tau) \right\| \\
\leq\ & \sup \left\| \hat{M}_n(\theta(\tau),\tau) - E\left[\hat{M}_n(\theta(\tau),\tau)\right] \right\| + \sup \left\| E\left[\hat{M}_n(\theta(\tau),\tau)\right] - M^u(\theta(\tau),\tau) \right\| \\
=\ & o_p(1) + o_p(1) \\
=\ & o_p(1) \qquad (.15)
\end{aligned}$$

So far, the stochastic convergence condition of Theorem 5.9 in van der Vaart (2000)[24] can be satisfied. (9) can be expressed as

$$\begin{aligned}
& M^u(\theta(\tau),\tau) \qquad (.16) \\
\equiv\ & E\left\{z_i\left[I\left\{\Lambda_i(Y,X,\theta(\tau)) \leq 0\right\} - \tau\right]\right\} \\
=\ & E\left(E\left\{z_i\left[I\left\{\Lambda_i(Y,X,\theta(\tau)) \leq 0\right\} - \tau\right] |z_i\right\}\right) \\
=\ & E\left\{z_i\left[P(\Lambda_i(Y,X,\theta(\tau)) \leq 0|z_i) - \tau\right]\right\} \qquad (.17)
\end{aligned}$$

Based on Assumption 2 and Assumption 9 , using the control convergence theorem

$$
\begin{aligned}
& \lim_{\delta\to 0} P(\Lambda_i(Y,X,\theta(\tau)+\delta) \le 0|z_i) && (.18)\\
= & \lim_{\delta\to 0} E(I\{\Lambda_i(Y,X,\theta(\tau)+\delta) \le 0\}|z_i)\\
= & E(\lim_{\delta\to 0} I\{\Lambda_i(Y,X,\theta(\tau)+\delta) \le 0\}|z_i)\\
= & E(I\{\Lambda_i(Y,X,\theta(\tau)) \le 0\}|z_i,\Lambda_i \neq 0)P(\Lambda_i \neq 0|z_i)\\
+ & E(\lim_{\delta\to 0} I\{\Lambda_i(Y,X,\theta(\tau)+\delta) \le 0\}|z_i,\Lambda_i = 0)P(\Lambda_i = 0|z_i)\\
= & E(I\{\Lambda_i(Y,X,\theta(\tau)) \le 0\}|z_i)\\
= & P(\Lambda_i(Y,X,\theta(\tau)) \le 0|z_i) && (.19)
\end{aligned}
$$

The continuity of $M^u(\theta(\tau),\tau)$ in the parameter space is proved, assuming that the moment function obtains the minimum value at $\tilde{\theta}(\tau)$, the following proof The sufficiency of the identification condition. Based on the assumption 2 , for any $\epsilon > 0$, $\|\theta(\tau)-\theta_0(\tau)\| \ge \epsilon$ holds, there is

$$
\inf \|M^u(\theta(\tau),\tau)\| = \min \|M^u(\theta(\tau),\tau)\| = \|M^u(\tilde{\theta}(\tau),\tau)\| > 0 \qquad (.20)
$$

Then, the consistency of parameter $\hat{\theta}_{\mathrm{MM}}$ estimation can be proved by Theorem 5.9 in van der Vaart (2000)[24].

Prove the consistency of the parameters $\hat{\theta}_{\mathrm{SGMM}}$. Based on the theorem 5.7 in van der Vaart (2000)[24], first prove that

$$
\sup \left\| \hat{M}_n(\theta(\tau),\tau)^T \hat{\Omega}\ \hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau)^T \Omega\ M^u(\theta(\tau),\tau) \right\| = o_p(1) \qquad (.21)
$$

Let $\|\cdot\|$ denote the F-norm: $\|A\| = \|A^T\| = \sqrt{tr(AA^T)}$. (.21) splits on the left

$$
\begin{aligned}
& \sup \left\| \hat{M}_n(\theta(\tau),\tau)^\top \hat{\Omega}\ \hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau)^\top \Omega\ M^u(\theta(\tau),\tau) \right\|\\
= & \sup \Big\| [\hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau) + M^u(\theta(\tau),\tau)]^\top\\
& \times[\hat{\Omega}-\Omega+\Omega]\ [\hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau) + M^u(\theta(\tau),\tau)] - M^u(\theta(\tau),\tau)^\top \Omega\ M^u(\theta(\tau),\tau) \Big\|\\
= & \sup \| M^u(\theta(\tau),\tau)^\top (\hat{\Omega}-\Omega)\ M^u(\theta(\tau),\tau) + M^u(\theta(\tau),\tau)^\top \Omega\ (\hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau))\\
& +(\hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau))^\top \Omega\ M^u(\theta(\tau),\tau)\\
& +M^u(\theta(\tau),\tau)^\top (\hat{\Omega}-\Omega)(\hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau))
\end{aligned}
$$

$$
\begin{aligned}
&+(\hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau))^\top \Omega\ (\hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau)) \\
&+(\hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau))^\top (\hat{\Omega} - \Omega)\ M^u(\theta(\tau),\tau) \\
&+(\hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau))^\top (\hat{\Omega} - \Omega)\ (\hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau))\| \qquad (.22)
\end{aligned}
$$

Combining triangle inequality and Cauchy inequality, the above equation satisfies

$$
\begin{aligned}
& \sup \left\| \hat{M}_n(\theta(\tau),\tau)^\top \hat{\Omega}\ \hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau)^\top \Omega\ M^u(\theta(\tau),\tau) \right\| \\
\leq\ & \sup \left\| M^u(\theta(\tau),\tau)^T \right\| \left\| \hat{\Omega} - \Omega \right\| \sup \left\| M^u(\theta(\tau),\tau) \right\| \\
+\ & \sup \left\| M^u(\theta(\tau),\tau)^\top \right\|\ \ \|\Omega\|\ \sup \left\| \hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau) \right\| \\
+\ & \sup \left\| (\hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau))^\top \right\|\ \ \|\Omega\|\ \sup \left\| M^u(\theta(\tau),\tau) \right\| \\
+\ & \sup \left\| M^u(\theta(\tau),\tau)^\top \right\| \left\| \hat{\Omega} - \Omega \right\| \sup \left\| \hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau) \right\| \\
+\ & \sup \left\| (\hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau))^\top \right\|\ \ \|\Omega\|\ \sup \left\| \hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau) \right\| \\
+\ & \sup \left\| (\hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau))^\top \right\| \left\| \hat{\Omega} - \Omega \right\| \sup \left\| M^u(\theta(\tau),\tau) \right\| \\
+\ & \sup \left\| (\hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau))^\top \right\| \left\| \hat{\Omega} - \Omega \right\| \sup \left\| \hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau) \right\| \qquad (.23)
\end{aligned}
$$

Based on (.15) and assuming 14

$$
\sup \left\| \hat{M}_n(\theta(\tau),\tau) - M^u(\theta(\tau),\tau) \right\| = o_p(1)
$$

$$
\hat{\Omega} - \Omega = o_p(1), \quad \|\Omega\| = O(1)
$$

and the continuity of the function $M^u(\theta(\tau),\tau)$, combined with Assumption 2

$$
\sup \|M^u(\theta(\tau),\tau)\| = O(1)
$$

Theorem 5.7 in van der Vaart (2000)[24] proves the sufficiency of the second condition. Since $M^u(\theta(\tau),\tau)^T \Omega\ M^u(\theta(\tau),\tau)$ is a continuous function of $\theta(\tau)$ and $\theta(\tau)$ is a compact set, let $\check{\theta}(\tau)$ represents the point corresponding to the minimum value of the function, then

$$
\begin{aligned}
& \inf M^u(\theta(\tau),\tau)^T \Omega\ M^u(\theta(\tau),\tau) \\
=\ & \min M^u(\theta(\tau),\tau)^T \Omega\ M^u(\theta(\tau),\tau) \\
=\ & M^u(\check{\theta}(\tau),\tau)^T \Omega\ M^u(\check{\theta}(\tau),\tau) \qquad (.24)
\end{aligned}
$$

Based on Assumption 2 and Assumption 14

$$M^u(\check{\theta}(\tau),\tau)^T\Omega\ M^u(\check{\theta}(\tau),\tau) > 0 \tag{.25}$$

So far, from the theorem 5.7 in van der Vaart (2000)[24], the consistency of parameters $\hat{\theta}_{\mathrm{SGMM}}$ (**??**) is proved.

It is worth mentioning that the consistency of the smoothed generalized moment estimator can also be proved by referring to theorem 3.4 in White (1996)[25] about the consistency of the extreme value estimator.

Appendix .3. Asymptotic Normality and Asymptotic Efficiency

The expectation of the sample moment condition can be expressed as

$$\begin{aligned} E\left[\hat{M}_n(\theta_0(\tau),\tau)\right] &= E\left[\frac{1}{n}\sum_{i=1}^{n} g_{ni}(\theta_0(\tau),\tau)\right] \\ &= E\left[g_{ni}(\theta_0(\tau),\tau)\right] \end{aligned} \tag{.26}$$

Based on the formula (.6)

$$\begin{aligned} E\left[\hat{M}_n(\theta_0(\tau),\tau)\right] &= \frac{h_n^r}{r!}\left[\int_{-1}^{1} \upsilon^r \widetilde{I}'(\upsilon)d\upsilon\right] E\left[Z_i f_{\Lambda|z}^{(r-1)}(0\mid Z_i)\right] + O(h_n^{r+1}) \\ &= O(h_n^r) \end{aligned} \tag{.27}$$

Combining Assumption 6, it is not difficult to get $\sqrt{n}E\left[\hat{M}_n(\theta_0(\tau),\tau)\right] = o_p(1)$. Based on assumption 13 , it can be proved that

$$\begin{aligned} \sqrt{n}\hat{M}_n(\theta_0(\tau),\tau) &= \sqrt{n}\hat{M}_n(\theta_0(\tau),\tau) - E\left[\hat{M}_n(\theta_0(\tau),\tau)\right] + \sqrt{n}E\left[\hat{M}_n(\theta_0(\tau),\tau)\right] \\ &= \sqrt{n}\hat{M}_n(\theta_0(\tau),\tau) - E\left[\hat{M}_n(\theta_0(\tau),\tau)\right] + o_p(1) \\ &\xrightarrow{d} N(0,\Sigma_\tau) \end{aligned} \tag{.28}$$

According to Assumption 12 , when the quantile assumption of the residual item is met, $\Sigma_\tau = \tau(1-\tau)E(z_i z_i^T)$.

In order to prove the large sample asymptotic normality of the estimated parameters, we first make some symbolic explanations

$$\nabla_{\theta^\top}\hat{M}_n(\theta_0(\tau),\tau) \equiv \frac{\partial}{\partial\theta^\top}\hat{M}_n(\theta_0(\tau),\tau)\ |_{\theta=\theta_0(\tau)} \tag{.29}$$

Use $\hat{M}_n^{(k)}$ to represent the k-th in the vector $\hat{M}_n^{(k)}(\theta(\tau),\tau)$ elements, define

$$\dot{M}_n(\tau) \equiv \left(\nabla_\theta \hat{M}_n^{(1)}(\tilde{\theta}_{(1)}(\tau),\tau), ..., \nabla_\theta \hat{M}_n^{(d_\theta)}(\tilde{\theta}_{(d_\theta)}(\tau),\tau)\right)^\top \tag{.30}$$

Since $\tilde{\theta}_{(k)} \in [\theta_0(\tau), \hat{\theta}_{\mathrm{SGMM}}(\tau)]$, based on the moment condition, Taylor series expansion at the true model parameter $\theta_0(\tau)$

$$\begin{aligned}
0 &= \left[\nabla_{\theta^\top} \hat{M}_n\left(\hat{\theta}_{\mathrm{SGMM}}(\tau),\tau\right)\right]^\top \hat{\Omega}\hat{M}_n\left(\hat{\theta}_{\mathrm{SGMM}}(\tau),\tau\right) \\
&= \left[\nabla_{\theta^\top} \hat{M}_n\left(\hat{\theta}_{\mathrm{SGMM}}(\tau),\tau\right)\right]^\top \hat{\Omega}\hat{M}_n\left(\theta_0(\tau),\tau\right) + \\
&\quad \left[\nabla_{\theta} \hat{M}_n\left(\hat{\theta}_{\mathrm{SGMM}}(\tau),\tau\right)\right]^\top \hat{\Omega}\dot{M}_n(\tau)\left(\hat{\theta}_{\mathrm{SGMM}}(\tau) - \theta_0(\tau)\right)
\end{aligned}$$

Based on (13) and Assumption 11 , using the continuous mapping theorem

$$[\dot{M}_n(\tau)]^{-1} \xrightarrow{p} G^{-1}$$

and

$$\hat{\Omega} \xrightarrow{p} \Omega, \quad [\nabla_{\theta^\top} \hat{M}_n(\hat{\theta}_{\mathrm{SGMM}}(\tau),\tau)]^\top \xrightarrow{p} G$$

Based on (.28) and Slutsky theorem

$$\begin{aligned}
& \sqrt{n}(\hat{\theta}_{\mathrm{SGMM}}(\tau) - \theta_0(\tau)) && (.31) \\
= \;& -[\nabla_{\theta^\top} \hat{M}_n(\hat{\theta}_{\mathrm{SGMM}}(\tau),\tau)]^\top \hat{\Omega}\ \dot{M}_n(\tau)^{-1} [\nabla_{\theta^\top} \hat{M}_n(\hat{\theta}_{\mathrm{SGMM}}(\tau),\tau)]^\top \hat{\Omega}\sqrt{n}\hat{M}_n(\theta_0(\tau),\tau) \\
= \;& -\{G^\top \Omega\ G\}^{-1} G^\top \Omega \sqrt{n}\hat{M}_n(\theta_0(\tau),\tau) + o_p(1) \\
\xrightarrow{d} \;& N(0, (G^\top \Omega\ G)^{-1} G^\top \Omega\ \Sigma_\tau\ \Omega\ G (G^\top \Omega\ G)^{-1}) && (.32)
\end{aligned}$$

Since the weight matrix $\Omega \equiv \Sigma_\tau^{-1}$ is asymptotically efficient, the asymptotic covariance matrix $(G^\top \Omega\, G)^{-1}$ and the SGMM covariance matrix estimator is semi-negative definite, which proves the asymptotic validity of the two-step estimator. In summary, the SGMM estimator has excellent large-sample properties.