Econometric Theory, **0**, 2026, 1–51.
doi:10.1017/S0266466626100346

ROBUST ESTIMATION FOR THE SPATIAL AUTOREGRESSIVE MODEL

Tuo Liu
Xiamen University

Xingbai Xu
Xiamen University

Lung-fei Lee
Shanghai University of Finance and Economics

Yingdan Mei
Renmin University of China

This article proposes and studies two Huber-type estimation approaches, namely, the Huber instrumental variable (IV) estimation and the Huber generalized method of moments (GMM) estimation, for a spatial autoregressive model. We establish the consistency, asymptotic distributions, finite sample breakdown points, and influence functions of these estimators. Simulation studies show that compared to the corresponding traditional estimators (the two-stage least squares estimator, the best IV estimator, and the GMM estimator), our estimators are more robust when the unknown disturbances are long-tailed, and our estimators only lose a little efficiency when the disturbances are short-tailed. Moreover, the Huber GMM estimator also outperforms several robust estimators in the literature. Finally, we apply our estimation method to investigate the impact of the urban heat island effect on housing prices. A package is published on GitHub for practitioners to use in their empirical studies.

1. MOTIVATION AND INTRODUCTION

In recent years, spatial econometric models have become increasingly popular in investigating spillover effects or competition in various fields of economics, including but not limited to tax competition (Lockwood and Migali, 2009; Lyytikäinen, 2012, Baskaran, 2014), the relationship between democracy and economic growth (Acemoglu et al., 2019), the impact of social networks on education (Calvó-Armengol, Patacchini, and Zenou, 2009; Liu and Lee, 2023), and international trade (Lee, Yang, and Yu, 2023). The spatial autoregressive (SAR) model is one of the most prevalent models in this area. Correspondingly,

We sincerely thank the Editor (Professor Peter C. B. Phillips), the Co-Editor (Professor Guido Kuersteiner), and four anonymous referees for their helpful comments and suggestions on the earlier versions of this article. Address correspondence to Xingbai Xu, Wang Yanan Institute for Studies in Economics (WISE), Department of Statistics and Data Science, School of Economics, Xiamen University, Xiamen, China, email: xuxingbai@xmu.edu.cn.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (https://creativecommons.org/licenses/by-nc-nd/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use and/or adaptation of the article.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

several estimation methods have been proposed and thoroughly studied for the SAR model, including quasi maximum likelihood estimation (QMLE), two-stage least squares (2SLS) estimation, the best instrumental variable (IV) estimation, the generalized method of moments (GMM), etc. The estimation methods mentioned above work well when the disturbances are short-tailed distributed, but are very sensitive to outliers or long-tailed disturbances. Our simulation studies reveal that these classical estimators tend to possess large variances in such circumstances. However, data with outliers or long-tailed disturbances are prevalent in economics and finance, e.g., housing prices, students' academic grades, income, yield rates of stocks or other financial assets, and the air quality index. As an example, this article employs an SAR model to investigate the housing prices in Beijing and documents that the estimated disturbances are long-tailed. This observation motivates us to develop some robust estimation methods for the SAR model that can better handle such scenarios.

This article is inspired by the Huber M-estimation proposed in Huber (1964), one of the mainstream robust estimation methods. An M-estimator is defined as $\arg\min_\theta \frac{1}{n}\sum_{i=1}^{n}\rho(x_i,\theta)$, where $\rho(\cdot)$ is a loss function. It is well known that least squares estimators are not robust to outliers and long-tailed disturbances for linear regression models. Huber (1964) combines the square loss function and the absolute value loss function to form a new loss function. M-estimators based on this loss function enjoy the merits of both least squares estimators and least absolute deviation (LAD) estimators, i.e., it is robust to outliers and more efficient than LAD estimators. The original Huber M-estimators are based on the assumption that regressors are exogenous. A few researchers manage to develop Huber-type estimators that accommodate endogeneity. Wagenvoort and Waldmann (2002) extend the idea of generalized M-estimators in Mallows (1975) to develop two robust estimators for a linear model in the presence of endogeneity, namely, the two-stage generalized M-estimator (2SGM) and robust GMM estimator, respectively.[1] Kim and Muller (2007) generalize 2SLS estimation to the two-stage Huber estimation for a linear regression model with endogenous regressors. Sølvsten (2020) proposes an estimator that is robust to outliers in a linear model with many IVs. All the works above deal with linear regression models with independent data.

Next, we provide a brief overview of the robust estimation literature pertaining to the SAR model and related network econometric models. To date, such literature remains limited. Genton (2003) calculates the breakdown point of the MLE for a pure SAR model (i.e., an SAR model without any $x_{i,n}$) whose spatial weights matrix is a lower triangular matrix, and the breakdown point of the least median of squares for the following SAR model, $Y_i = \frac{\rho}{2}(Y_{i-1} + Y_{i+1}) + \epsilon_i$, where ϵ_i's are independently and identically distributed (i.i.d.) normal random variables. Britos and Ojeda (2019) propose a robust estimation procedure for an SAR model in

[1] The basic idea of Wagenvoort and Waldmann (2002) is to give outliers small weights (or even zero weights) in the IV equation, such that the influence of outliers is small.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

which individuals are located on a two-dimensional integer lattice $\mathbb{Z}^2$. They adopt a lower triangular spatial weights matrix which precludes endogeneity. Yildirim and Mert Kantar (2020) study the robust estimation of a spatial error model by adjusting the score of the log-likelihood function. They focus on computational aspects and simulation studies of the method. Xiao, Xu, and Zhong (2023) study the Huber estimation of the network autoregressive model. But neither the model in Britos and Ojeda (2019) nor that in Xiao, Xu, and Zhong (2023) involves an endogenous spatial lag term, distinguishing their models from the SAR model. The following two papers are more related to our work. Su and Yang (2011) propose the IV quantile regression (IVQR) approach, which is robust and less sensitive to outliers. However, we find that the IVQR approach suffers from significant efficiency loss when disturbances are short-tailed distributed. Tho et al. (2023) study the Huber estimation of the SAR model. They adapt the score of the log-likelihood function, which relies on i.i.d. normal assumption, to develop an estimator that is robust to outliers.

Unlike Tho et al. (2023), our work generalizes 2SLS and GMM estimation for the SAR model to ensure robustness against outliers, long-tailed disturbances, and conditional heteroskedasticity, without sacrificing much efficiency under short-tailed disturbances. We integrate the idea of the Huber M-estimation with traditional moment estimation methods for the SAR model (the 2SLS, the best IV [BIV] estimation, and GMM) to propose two novel estimation approaches for the SAR model, i.e., Huber IV (HIV) estimation and Huber GMM (HGMM) estimation.

Traditional estimation methods (MLE and GMM) for the SAR model typically require the existence of moments of order higher than 4 for the error terms. The asymptotic normality of our proposed estimators only requires the existence of the first-order moment, without imposing any additional restrictive assumptions. Compared to the HIV estimation, the HGMM estimation is more recommended in empirical studies, as it is more efficient. Compared to GMM (or QML) estimation, the HGMM estimation is robust to outliers, long-tailed disturbances (and conditional heteroskedasticity). Even if the error terms are i.i.d. normally distributed, the root mean square error (RMSE) of the HGMM estimator is only about 0.2%–1.5% greater than that of the MLE, as demonstrated by simulation studies. In sum, when the error terms are i.i.d. and follow a short-tailed distribution, QML, GMM, and HGMM estimation all work satisfactorily. However, when outliers are present in the data or when it is uncertain whether the data follow a long-tailed distribution or not, the HGMM estimation method should be the preferred choice over traditional estimation methods. Practitioners can access the HGMM estimation via the accompanying package "HGMM_SAR" in their empirical studies.[2]

The rest of this article is organized as follows. In Section 2, we propose the HIV estimation and the HGMM estimation for the SAR model. In Section 3, we

[2] https://github.com/Xingbai2004/Huber-GMM

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

study the consistency and asymptotic distributions of these estimators and discuss their finite sample breakdown points and influence functions. In Section 4, we implement Monte Carlo simulations to assess the finite sample properties of our proposed estimators in comparison with traditional counterparts and some other robust estimators. Section 5 applies the HGMM estimation to examine the impact of the urban heat island effect on housing prices in Beijing. Section 6 concludes this article. Additional extensions of our methods, all proofs, and ancillary tables for simulation and empirical results are collected in appendixes.

Notations: $\mathbb{N} \equiv \{1, 2, 3, \ldots\}$ denotes the set of positive integers. Let $\mathbb{R}^n$ denote the n-dimensional euclidean space. For any $m \times n$ matrix $A = (a_{ij})$, $A_{i\cdot}$ denotes its i-th row, $\|A\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{m} |a_{ij}|$, $\|A\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|$, and its Frobenius norm $||A|| \equiv (\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2)^{1/2}$. For any square matrix $A = (a_{ij})$, $\mathrm{tr}(A)$ denotes its trace. For any random vector x and constant $p \ge 1$, the L_p-norm of X is defined to be $||x||_{L^p} \equiv (\mathrm{E}\,||x||^p)^{1/p}$. For any matrix A, $\min \mathrm{sv}(A)$ and $\max \mathrm{sv}(A)$ denote its minimum and maximum singular values, respectively. For any set A, $|A|$ denotes its cardinality, i.e., the number of elements in A. Let $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and distribution, respectively. $f(x) \lesssim g(x)$ means that $f(x) \le Cg(x)$ for all x for some constant C. For any scalar-valued function $f : X(\subset \mathbb{R}) \to \mathbb{R}$ and any vector $a = (a_1, \ldots, a_n)'$ with each $a_i \in X$, we extend the definition of f to $f(a) = (f(a_1), \ldots, f(a_n))'$. A random field about $x_{i,n}$ is denoted by $\{x_{i,n} : i \in D_n, n \in \mathbb{N}\}$, where D_n denotes the set of individuals,[3] and sometimes $\{x_{i,n} : i \in D_n\}$ or $\{x_{i,n}\}$ for short.

2. TWO ROBUST ESTIMATORS FOR THE SAR MODEL

In this section, we propose two robust estimators for the SAR model based on Huber's loss function: the HIV estimator and the HGMM estimator. We leave the discussion of their theoretical properties to Section 3.

2.1. Huber IV Estimator of the SAR Model

Kelejian and Prucha (1998) propose the 2SLS estimator for the SAR model. Building upon their approach, we integrate Huber's (1964) robust estimation technique to propose the HIV estimator in this section. The SAR model is formulated as

$$y_{i,n} = \lambda w_{i\cdot,n} Y_n + x_{i,n}'\beta + \epsilon_{i,n} \equiv \lambda \sum_{j=1}^{n} w_{ij,n} y_{j,n} + x_{i,n}'\beta + \epsilon_{i,n}, \tag{2.1}$$

or in matrix form, $Y_n = \lambda W_n Y_n + X_n\beta + \epsilon_n$, where $Y_n \equiv (y_{1,n}, \ldots, y_{n,n})'$ is the column vector of dependent variables, $W_n = (w_{ij,n})$ is an $n \times n$ non-stochastic spatial weights matrix and $w_{i\cdot,n}$ is its i^{th} row, $x_{i,n} = (x_{i1}, \ldots, x_{ip_x})'$ is a p_x-dimensional column vector of exogenous regressors, including the intercept term

[3] See p.7 for a detailed explanation.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

$x_{i1} = 1$, $X_n = (x_{1,n}, \ldots, x_{n,n})'$, and $\epsilon_n = (\epsilon_{1,n}, \ldots, \epsilon_{n,n})'$ is the column vector of disturbances. Denote the parameter vector $\theta = (\lambda, \beta')'$ and the true parameter vector $\theta_0 = (\lambda_0, \beta_0')'$.

Huber (1964) introduces the following loss function:

$$\rho(t) = \begin{cases} \frac{1}{2}t^2 & |t| \le M \\ M|t| - \frac{1}{2}M^2 & |t| > M, \end{cases}$$

where M is a critical value. When $M = \infty$, $\rho(\cdot)$ is the squared loss function, while as $M \to 0$, $\rho(\cdot)$ approximates the absolute value loss function. Since $w_{i\cdot,n}Y_n$ is an endogenous variable, the standard Huber M-estimator based on the loss function $\rho(\cdot)$ is typically inconsistent. However, the idea of moment estimation based on the derivative of $\rho(\cdot)$ can be adapted to the SAR model, provided that suitable IVs are available.

Notice that $\rho(t)$ is a continuously differentiable even function, and its derivative is

$$\psi(t) = \frac{d\rho(t)}{dt} = \begin{cases} M & t > M \\ t & |t| \le M \\ -M & t < -M. \end{cases} \tag{2.2}$$

Suppose that

$$\mathrm{E}[\psi(\epsilon_{i,n})|X_n] = 0, \tag{2.3}$$

which is a standard assumption in the Huber estimation literature, and $\epsilon_{i,n}$'s are independent conditional on X_n. When the distribution of $\epsilon_{i,n}$ is symmetric about 0 given $x_{i,n}$, Eq. (2.3) holds trivially. When $\epsilon_{i,n}$'s are i.i.d. but not symmetrically distributed conditional on X_n, we can also shift the intercept such that Eq. (2.3) is satisfied,[4] and that is why an intercept term must be included. Notice that Eq. (2.3) allows conditional heteroskedasticity, so $\epsilon_{i,n}$ and $x_{i,n}$ might not be independent.

Since W_n is nonstochastic, Eq. (2.3) implies

$$\mathrm{E}[\psi(\epsilon_{i,n})x_{i,n}] = 0 \text{ and } \mathrm{E}[\psi(\epsilon_{i,n})w_{i\cdot,n}X_n] = 0.$$

Hence, $x_{i,n}$ and $w_{i\cdot,n}X_n$ can be regarded as IVs.[5] We can also add the second-order spatial effects $(w_{i\cdot,n}W_nX_n)$ into the IV vector $q_{i,n}$, or use $G_{i\cdot,n}X_n\beta_0$ as an IV for the endogenous variable $w_{i\cdot,n}Y_n$ in the SAR model, where $G_n \equiv W_n(I_n - \lambda_0 W_n)^{-1}$

[4] Suppose that conditional on X_n, $\epsilon_{i,n}$'s are the i.i.d. copies of a continuous random variable ϵ with a continuous density function $f(\epsilon)$ supported on $\mathbb{R}$ and $F(\cdot)$ denotes its distribution function. By Lemma A.7(2), whenever $\Pr(-M - a \le \epsilon \le M - a) > 0$ for all a, $\frac{\partial}{\partial a}\mathrm{E}[\psi(\epsilon + a)] = F(M + a) - F(-M + a) > 0$. So, $\xi(a) \equiv \mathrm{E}[\psi(\epsilon + a)]$ is a strictly increasing and continuous function on $\mathbb{R}$. In addition, $\lim_{a\to-\infty}\xi(a) = -M$ and $\lim_{a\to+\infty}\xi(a) = M$. Therefore, from the intermediate value theorem, the equation $\xi(a) = 0$ must have a unique solution a_0. In other words, we can guarantee Eq. (2.3) to hold by shifting the intercept.

[5] If W_n is row-normalized and the first column of X_n is the intercept term $\iota_n \equiv (1, \ldots, 1)'$, then the first element of $w_{i\cdot,n}X_n$ equals 1. In this case, we have to delete the first column in $w_{i\cdot,n}X_n$ to eliminate complete multicollinearity among IVs.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

and $G_{i\cdot,n}$ is its i-th row.[6] We can also censor these IVs such that they can satisfy the theory. Let $q_{i,n} \in \mathbb{R}^{p_q}$ be the column IV vector in the following and $Q_n \equiv (q_{1,n}, \ldots, q_{n,n})'$ denote the corresponding $n \times p_q$ IV matrix, where p_q must be greater than or equal to $p_x + 1$, the dimension of θ. Then, we can write the moment conditions as

$$\mathrm{E}[\psi(\epsilon_{i,n})q_{i,n}] = 0. \tag{2.4}$$

Based on Eq. (2.4) and the corresponding sample moments, we can derive an estimator for θ, which we term the HIV estimator. Denote $\epsilon_{i,n}(\theta) \equiv y_{i,n} - \lambda w_{i\cdot,n} Y_n - x'_{i,n}\beta$, $\epsilon_n(\theta) \equiv (\epsilon_{1,n}(\theta), \ldots, \epsilon_{n,n}(\theta))'$, and $\psi(\epsilon_n(\theta)) = (\psi(\epsilon_{1,n}(\theta)), \ldots, \psi(\epsilon_{n,n}(\theta)))'$. The sample moments corresponding to Eq. (2.4) are $\frac{1}{n}Q'_n\psi(\epsilon_n(\theta)) = \frac{1}{n}\sum_{i=1}^n \psi(\epsilon_{i,n}(\theta))q_{i,n}$. Given a positive-definite GMM weighting matrix Ω_n, we can obtain an HIV estimator $\hat{\theta}_n = \arg\min_{\theta\in\Theta} \hat{T}_n(\theta)$, where

$$\begin{aligned}\hat{T}_n(\theta) &= \left[\frac{1}{n}Q'_n\psi(\epsilon_n(\theta))\right]' \Omega_n \left[\frac{1}{n}Q'_n\psi(\epsilon_n(\theta))\right] \\ &= \left[\frac{1}{n}\sum_{i=1}^n \psi(\epsilon_{i,n}(\theta))q_{i,n}\right]' \Omega_n \left[\frac{1}{n}\sum_{i=1}^n \psi(\epsilon_{i,n}(\theta))q_{i,n}\right].\end{aligned} \tag{2.5}$$

In practice, we can let $\Omega_n \equiv \left(\frac{1}{n}\sum_{i=1}^n q_{i,n}q'_{i,n}\right)^{-1}$. In this case, when $M = \infty$, the HIV estimator is reduced to the 2SLS estimator based on $q_{i,n}$.

If conditional on X_n, $\epsilon_{i,n}$'s are i.i.d., the best HIV (BHIV) for $w_{i\cdot,n}Y_n$ is $G_{i\cdot,n}X_n\beta_0$, which is the same as that proposed in Lee (2003). To implement the BHIV estimation, we need a consistent estimator for θ_0 and replace G_n and β_0 by their estimates (see Section B of the Supplementary Material for details).

2.2. Huber GMM Estimator

Building upon the 2SLS in Kelejian and Prucha (1998), the quadratic moment conditions for a pure SAR model in Kelejian and Prucha (1999), and the BIV estimation in Lee (2003) for the SAR model, Lee (2007) proposes a GMM estimator for an SAR model that employs both linear and quadratic moments. This GMM estimator surpasses the 2SLS estimator in efficiency and can rival the MLE when the disturbances are normally distributed and appropriate quadratic moments are chosen. Lin and Lee (2010) extend Lee (2007) to allow heteroskedasticity.

Like a 2SLS estimator, the HIV estimator proposed in the previous section employs only the first-order moments based on the conditional IV moments motivated by Eq. (2.3). To enhance estimation efficiency, drawing inspiration from Kelejian and Prucha (1999), Lee (2007), and Lin and Lee (2010), we explore "quadratic Huber moments" in this section.

[6] In practice, as λ_0 and β_0 are unknown, we need initial consistent estimates for them and use the corresponding feasible IV in estimation.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

Our discussion proceeds under the same assumptions outlined in Section 2.1, especially Eq. (2.3). Let $P_{kn} = (P_{ij,kn})$ $(k = 1, \ldots, m)$ be n-dimensional nonstochastic square matrices with zero diagonals, i.e., $P_{ii,kn} = 0$ for all $i = 1, \ldots, n$. Then, it follows from Eq. (2.3) and the independence of $\epsilon_{i,n}$'s that

$$\begin{aligned}\mathrm{E}[\psi(\epsilon_n)' P_{kn} \psi(\epsilon_n)] &= \sum_{i=1}^{n} \sum_{j \neq i} P_{ij,kn} \, \mathrm{E}[\psi(\epsilon_{i,n}) \psi(\epsilon_{j,n})] \\ &= \sum_{i=1}^{n} \sum_{j \neq i} P_{ij,kn} \, \mathrm{E}[\psi(\epsilon_{i,n})] \, \mathrm{E}\, \psi(\epsilon_{j,n})] = 0.\end{aligned}$$

Consequently, we can construct m sample quadratic Huber moments: $\frac{1}{n}\psi(\epsilon_n(\theta))'$ $P_{kn}\psi(\epsilon_n(\theta))$. As in Section 2.1, $q_{i,n}$ denotes a column IV vector and $Q_n \equiv (q_{1,n}, \ldots, q_{n,n})'$. We construct a sample moment vector as in Lee (2007) and Lin and Lee (2010):

$$g_n(\theta) = \frac{1}{n} \begin{pmatrix} \psi(\epsilon_n(\theta))' P_{1n} \psi(\epsilon_n(\theta)) \\ \vdots \\ \psi(\epsilon_n(\theta))' P_{mn} \psi(\epsilon_n(\theta)) \\ Q_n' \psi(\epsilon_n(\theta)) \end{pmatrix}. \tag{2.6}$$

Given a GMM weighting matrix Ω_n, we obtain an HGMM estimator

$$\hat{\theta}_{HGMM} = \arg\min_{\theta} g_n(\theta)' \Omega_n g_n(\theta).$$

2.3. Selection of the Critical Value *M*

The Huber loss function contains a critical value M. In Section 3, we show that the two proposed estimators are consistent and asymptotically normal for any fixed constant M. Nonetheless, in finite samples, it might be beneficial to employ a data-driven approach to select a suitable M that balances robustness and efficiency. Following Bickel (1975), we propose to use

$$M = c \cdot \text{median}\left(|\epsilon - \text{median}(\epsilon)|\right)/0.6745, \tag{2.7}$$

where c is a positive constant and $\text{median}\left(|\epsilon - \text{median}(\epsilon)|\right)/0.6745$ serves as a robust estimator for the standard deviation. The factor 0.6745 corresponds to the 75% quantile of the standard normal distribution.

Huber (1964) and Maronna et al. (2019) demonstrate through simulation studies that the performance of Huber M estimators is relatively insensitive to the choice of c, when it lies between 1 and 1.5. Any value within the interval $[1, 1.5]$ can yield satisfactory results. Both MATLAB and R default to a critical value of 1.345, which corresponds to a 95% relative asymptotic efficiency for location estimators under normality.[7] We also investigate this issue through simulations in Section 4.

[7] We are grateful to an anonymous referee for pointing this out. Fan et al. (2020, p.298) also mention this.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

The conclusion is similar. When c is a constant between 1 and 1.5, our proposed estimators outperform traditional estimators for the SAR model under long-tailed disturbances or extreme outliers, with performances being comparable across this range. Thus, it makes little practical difference whether $c = 1$, 1.2, 1.345, or 1.5. Following the practice in MATLAB and R, we determine the critical value M using Eq. (2.7) with $c = 1.345$ in both our simulation studies and empirical applications.

3. THEORETICAL ANALYSIS

In this section, we will establish the large sample properties of the HIV estimator and the HGMM estimator, and discuss their finite sample breakdown points and influence functions.

3.1. Spatial Near-Epoch Dependence

Since Kelejian and Prucha (2001) establish a central limit theorem (CLT) for linear-quadratic forms in the context of SAR models, the theory of linear-quadratic forms has become a cornerstone in spatial econometrics (see Kelejian and Prucha, 2001; Lee, 2004, 2007). However, as $\psi(\epsilon_{i,n}(\theta))$ does not conform to a linear-quadratic form, the theory of linear-quadratic forms is not applicable to Huber estimators, except when $\theta = \theta_0$. Hence, we require certain weak spatial dependence concepts such that some laws of large numbers (LLN) can be utilized to establish asymptotic properties of our estimators. The theory of the spatial near-epoch dependence (NED) developed in Jenish and Prucha (2012) is prevalent in spatial econometrics and is capable of handling nonlinear models. Thus, using this approach, we establish large sample properties of our estimators via spatial NED theory.

Suppose that there are n individuals located in $D_n \subset \mathbb{R}^d$ and $|D_n| = n$. As in Jenish and Prucha (2012), the distance between any two points $x, y \in \mathbb{R}^d$ is defined as $d_{xy} = \max_{i=1,\ldots,d} |x_i - y_i|$. We index these n individuals as $1, 2, \ldots, n$. We identify their indices and locations in this article.[8]

Next, we list the assumptions required to establish some moment conditions and spatial NED properties for $\{y_{i,n} : i \in D_n\}$ and related variables, which will be applied to establish the asymptotic theories in Sections 3.2 and 3.3.

Assumption 1. For any two individuals i and j in D_n, their distance $d_{ij} \geq 1$.

Assumption 2. (i) $\epsilon_{i,n}$'s are independent over i, and $\sup_{i,k,n} \left(\left\| x_{ik,n} \right\|_{L^\gamma} + \left\| \epsilon_{i,n} \right\|_{L^\gamma} \right) \leq C_{x\epsilon} < \infty$ for some constants $\gamma \in [1, \infty]$.

[8] To be clear, we present a concrete example. Suppose $n = 3$ individuals are located in $\mathbb{R}^2$, their locations are $(x_1, y_1)' = (0,0)'$, $(x_2, y_2)' = (1,0)'$, and $(x_3, y_3)' = (3, 4.4)'$. Then, $D_n = \{(0,0)', (1,0)', (3,4.4)'\}$. For the third individual, we will use either $i = 3$ or $i = (3.3, 4.4)'$ to represent it. And the distance between individuals 1 and 3 is $d_{13} = \max(|3-0|, |4.4-0|) = 4.4$, rather than $3 - 1 = 2$.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

(ii) $\{x_{i,n}\}_{i=1}^{n}$ is an α-mixing random field with α-mixing coefficient $\alpha(u,v,r) \leq C_{x1}(u+v)^{\tau}\exp(-C_{x2}r^{C_{x3}})$ for some constants $\tau \geqslant 0$ and $C_{x1}, C_{x2}, C_{x3} > 0$.[9]
(iii) $\mathrm{E}[\psi(\epsilon_{i,n})|X_n] = 0$.

Assumption 3. $\zeta \equiv \lambda_m \sup_n ||W_n||_\infty < 1$, where $\Lambda = [-\lambda_m, \lambda_m]$ represents the parameter space of λ on $\mathbb{R}$ and λ_m is the upper bound of Λ.

Assumption 4. In addition to $w_{ii,n} = 0$ for all i and n, all the weights $w_{ij,n}$'s in the non-stochastic and nonzero $n \times n$ matrix W_n satisfy at least one of the following two conditions:
(i) When $d_{ij} > \bar{d}_0$, where $\bar{d}_0 > 1$ is a constant, $w_{ij,n} = 0$.
(ii) There exists an $\alpha > 2.5d$ and a constant C_0 such that $|w_{ij,n}| \leq C_0/d_{ij}^{\alpha}$.

Under Assumption 1, we consider the scenario of increasing domain asymptotics. This setting is widely presumed in spatial econometrics (Qu and Lee, 2015; Xu and Lee, 2015a, 2015b, 2018; Liu and Lee, 2019), as it is the foundation of the spatial mixing and NED theory (Jenish and Prucha, 2009, 2012). This setting is particularly intuitive when individuals are situated in geographical locations like countries and cities. Nonetheless, social and economic characteristics can also be employed to define coordinates, thereby constructing distances between entities. Although Assumption 1 is popular in spatial econometrics, it rules out the case of infilled asymptotics, i.e., within a bounded space, the number of individuals tends to infinity. Infill asymptotics is popular in time series, especially functional time series (see, e.g., Phillips and Jiang, 2025, among many other papers). In spatial statistics, it is also widely studied, e.g., the estimation of a Gaussian random field (see Cressie, 1993 and Gelfand et al., 2010 for details). Under infilled asymptotics, the spatial correlations of many observations might be strong and will not tend to zero. In this case, traditional estimation methods for the SAR model are generally inconsistent.[10]

Assumption 2(i) imposes some moment conditions on the exogenous variable $x_{i,n}$ and the error term $\epsilon_{i,n}$. Specifically, we require only the γ-th-order moment for $x_{i,n}$ and $\mathrm{E}|\epsilon_{i,n}|^{\gamma}$ for some $\gamma \geq 1$. For the QMLE and the GMM estimator, when exogenous variables are uniformly bounded, Lee (2004, 2007) requires disturbances to have a finite moment of order higher than 4, i.e., $\gamma > 4$ in our notation. Assumption 2(ii) allows $x_{i,n}$ to be spatially correlated. Assumption 2(iii) has been discussed below Eq. (2.3).

Assumption 3 guarantees that the von Neumann series expansion for $(I_n - \lambda W_n)^{-1}$ exists. This assumption is one of the critical assumptions leading

[9]Let $X \equiv \{X_{i,n}; i \in D_n, n \in \mathbb{N}\}$ be a triangular array of random vectors on a probability space $(\Omega, \mathcal{F}, \Pr)$. For two sub-$\sigma$-fields $\mathcal{A}$ and $\mathcal{B}$, the α-mixing coefficient between $\mathcal{A}$ and $\mathcal{B}$ is defined as $\alpha(\mathcal{A},\mathcal{B}) \equiv \sup(|\Pr(A \cap B) - \Pr(A)\Pr(B)| : A \in \mathcal{A}, B \in \mathcal{B})$. For $Q_1, Q_2 \subseteq D_n$, denote $\sigma_n(Q_1) \equiv \sigma(X_{i,n} : i \in Q_1)$ the σ-field generated by the random vectors $X_{i,n}$ with i in Q_1, and $\alpha_n(Q_1,Q_2) \equiv \alpha(\sigma_n(Q_1), \sigma_n(Q_2))$. The α-mixing coefficient for X is $\alpha_n(u,v,r) \equiv \sup_{Q_1,Q_2 \subseteq D_n}\{\alpha_n(Q_1,Q_2) : |Q_1| \leq u, |Q_2| \leq v, d(Q_1,Q_2) \geqslant r\}$, where $d(Q_1,Q_2) \equiv \inf\{d_{ij} : i \in Q_1, j \in Q_2\}$ is the distance between Q_1 and Q_2. Denote $\alpha(u,v,r) \equiv \sup_n \alpha_n(u,v,r)$.

[10]We thank the Editor, Professor Peter Phillips, for the comments about infilled asymptotics.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

to the NED of $\{y_{i,n}\}$ and $\{w_{i\cdot,n}Y_n\}$. It is also widely used in the spatial econometrics literature (see, e.g., the conditions in Lemma 2 in Kelejian and Prucha, 2010).

Assumption 4 allows two setups for the spatial weights matrix W_n. Under Assumption 4(i), two individuals have direct interactions only if they are within a fixed finite distance. Under Assumption 4(ii), two individuals are allowed to have direct interactions even if they are far away from each other, but the strength of their interactions should decrease as distance increases. Assumptions 1 and 4 imply that $\sup_n ||W_n||_\infty < \infty$ and $\sup_n ||W_n||_1 < \infty$. With Assumption 3, they further imply that $\sup_n ||(I_n - \lambda W_n)^{-1}||_\infty < \infty$. The uniform boundedness of the norms of W_n and $(I_n - \lambda W_n)^{-1}$ is prevalent in the literature of spatial econometrics (see, e.g., Assumption 3 in Kelejian and Prucha, 2010 and Assumption 5 in Lee, 2004).

Assumptions 1–4 imply that $\{y_{i,n}\}$ and $\{w_{i\cdot,n}Y_n\}$ are L_γ-NED on $\{x_{i,n}, \epsilon_{i,n} : i \in D_n\}$ (Lemma A.8). We summarize the moment and L_γ-NED properties of the regressor $z_{i,n} \equiv (w_{i\cdot,n}Y_n, x_{i,n}')'$, $\psi(\epsilon_{i,n}(\theta))$, and common choices for IVs in Lemma A.8 in Appendix A.

3.2. Large Sample Properties of the Huber IV Estimator

In this section, we establish the large sample properties of the HIV estimator defined in Eq. (2.5).

Assumption 5. The parameter space Θ is convex and compact.

Assumption 6. The IV vector $q_{i,n} \in \mathbb{R}^{p_q}$ $(p_q \geq p_x + 1)$ is constructed from X_n and W_n, and $\sup_{i,n} ||q_{i,n}|| < \infty$.

Assumption 7. $\Omega_n \xrightarrow{p} \Omega$ for some positive-definite matrix Ω.

Assumption 8. $\liminf_{n\to\infty} \left\| \frac{1}{n} \sum_{i=1}^n \mathrm{E}\left[\psi(\epsilon_{i,n}(\theta))q_{i,n}\right] \right\| > 0$ for any $\theta \neq \theta_0$.

In this article, we can use $w_{i\cdot,n}X_n$, $w_{i\cdot,n}W_nX_n$, or $G_{i\cdot,n}X_n\beta_0$ as the IV for $w_{i\cdot,n}Y_n$. When $x_{i,n}$'s are uniformly bounded, Assumption 6 holds for $x_{i,n}$, $w_{i\cdot,n}X_n$, $w_{i\cdot,n}W_nX_n$, and $G_{i\cdot,n}X_n\beta_0$; if $x_{i,n}$'s are unbounded, we may censor them such that $q_{i,n}$ is uniformly bounded. We assume $\{q_{i,n} : i \in D_n\}$ to be uniformly bounded to simplify the proof.[11] Alternatively, a more involved proof would permit unbounded $q_{i,n}$.

Assumption 7 is the standard requirement for a GMM weighting matrix.

Assumption 8 is crucial for ensuring the identifiable uniqueness of θ_0. It implies that θ_0 is the unique solution to $\frac{1}{n} \sum_{i=1}^n \mathrm{E}\left[\psi(\epsilon_{i,n}(\theta))q_{i,n}\right] = 0$ when n is large enough. However, Assumption 8 is slightly stronger than the condition that θ_0 is the unique solution to $\frac{1}{n} \sum_{i=1}^n \mathrm{E}\left[\psi(\epsilon_{i,n}(\theta))q_{i,n}\right] = 0$. Assumption 8 excludes the possibility that $0 < \left\| \frac{1}{n} \sum_{i=1}^n \mathrm{E}\left[\psi(\epsilon_{i,n}(\theta))q_{i,n}\right] \right\| \to 0$ as $n \to \infty$.

Under the above assumptions, the HIV estimator is consistent, as stated in Theorem 1.

[11] We thank the co-editor for this suggestion.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

THEOREM 1. *Under Assumptions 1–8, the HIV estimator* $\hat{\theta} \xrightarrow{p} \theta_0$.

Remark. Tho et al. (2023) have shown that their estimator is Fisher-consistent (Theorem 1 in their paper), i.e., $\mathrm{E}\,\eta(\theta_0, y) = 0$, where $\eta(\theta_0, y)$ is the modified score function in their notation. Fisher-consistency only cares about the behavior of moment functions at the true parameter θ_0 and $\psi(\epsilon_{i,n}(\theta_0)) = \psi(\epsilon_{i,n})$ are i.i.d. over i and n. As a result, the LLN for independent variables is applicable to $\psi(\epsilon_{i,n})x_{i,n}$ and the LLN for linear-quadratic forms is applicable to $\psi(\epsilon_{i,n})w_{i\cdot,n}Y_n$.

The consistency established in Theorem 1 is different from Fisher-consistency: the former usually requires stronger conditions. The Fisher-consistency is similar to the correct model specification condition for GMM estimators. In addition to the correct model specification condition, we need both the uniform convergence in probability condition and identifiable uniqueness condition for consistency. Consequently, we must also address the behavior of moments evaluated at $\theta \neq \theta_0$. When $\theta \neq \theta_0$, $\psi(\epsilon_{i,n}(\theta))q_{i,n}$ is no longer a linear-quadratic form of $\epsilon_{i,n}$'s and $x_{i,n}$'s. As a result, the theory of linear-quadratic forms is not applicable. Hence, we employ the spatial NED theory in our article.

Next, we derive the asymptotic distribution of the HIV estimator $\hat{\theta}$. Given that the objective function $\hat{T}_n(\theta)$ is not differentiable with respect to θ, the conventional Taylor expansion cannot be employed to derive the asymptotic distribution of $\hat{\theta}$. However, we can utilize the empirical process theory for derivation. We generalize Theorem 3.3 in Pakes and Pollard (1989), which studies the asymptotic distributions of GMM estimators under non-differentiable moment conditions, to accommodate heterogeneous data in Proposition B.2. The critical condition in Proposition B.2 is a condition similar to stochastic equicontinuity (SEC), specifically condition (iii) there. Within the empirical process literature, several results exist to validate the SEC condition of an empirical process. However, most of them are tailored to independent or stationary data. Data generated by spatial econometric models are usually both heterogeneous and spatially correlated. Therefore, we need to generalize those results to allow for both heterogeneity and spatial correlation. Pollard (1985, p. 305) extends the result in Huber (1967) to establish a set of sufficient conditions for SEC in the context of independent data based on a bracketing condition. In this article, we generalize the result in Pollard (1985) to account for serial or spatial correlation (see Proposition B.1 for details).

To establish the asymptotic distribution of $\hat{\theta}$, we need some additional regularity conditions. $\dot{\psi}(\epsilon) \equiv 1(|\epsilon| \leq M)$ equals the derivative of $\psi(\epsilon)$ almost everywhere. Define $\Gamma_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[\dot{\psi}(y_{i,n} - z'_{i,n}\theta)q_{i,n}z'_{i,n}]$, $\Gamma_n = \Gamma_n(\theta_0)$, and $V_n \equiv \mathrm{var}\left\{\frac{1}{n}\sum_{i=1}^{n} \psi(\epsilon_{i,n})q_{i,n}\right\}$. By the independence of $\epsilon_{i,n}$'s and the law of iterated expectation, straightforward calculations yield that

$$V_n = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[\psi(\epsilon_{i,n})^2 q_{i,n}q'_{i,n}].$$

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

Assumption 9. $\theta_0 \in \Theta^0$, where Θ^0 denotes the interior of the parameter space Θ.

Assumption 10. (i) The class of functions $\{\Gamma_n(\theta) : n \in \mathbb{N}\}$ is equicontinuous at θ_0; (ii) there exists a positive constant δ_0 such that

$$\liminf_{n\to\infty} \inf_{\theta_1,\ldots,\theta_l \in B(\theta_0,\delta_0)} \min \mathrm{sv} \left(\begin{bmatrix} \Gamma_{1n}(\theta_1) \\ \vdots \\ \Gamma_{ln}(\theta_l) \end{bmatrix} \right) > 0.$$

$\Gamma_{in}(\theta)$ is the i^{th} row of $\Gamma_n(\theta)$; and (iii) $\lim_{n\to\infty} \Gamma_n = \Gamma_0$ and Γ_0 has a full column rank.

Assumption 11. $\lim_{n\to\infty} V_n = V_0$ for some positive-definite matrix V_0.

Assumption 9 is a standard condition for establishing asymptotic distributions. Assumption 10 imposes some conditions on $\Gamma_n(\theta)$ and Γ_n such that condition (ii) of Proposition B.2 holds. We need different θ_i for different rows because there does not exist a mean-value theorem for vector-valued functions. Under standard conditions, $\Gamma_n(\theta)$ is continuous at θ_0. The equicontinuity at θ_0 for $\Gamma_n(\theta)$ means that as the sample size n increases to infinity, $\Gamma_n(\theta)$ does not change too much around θ_0. This is a reasonable requirement; otherwise, it would be infeasible to conduct any reliable inference based on finite samples.

THEOREM 2. *Under Assumptions 1–11, if the α in Assumption 4 satisfies* $\alpha > \left[\frac{2}{\min(2,\gamma)} + 1\right] d$,

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, (\Gamma_0'\Omega\Gamma_0)^{-1}\Gamma_0'\Omega V_0 \Omega\Gamma_0(\Gamma_0'\Omega\Gamma_0)^{-1}).$$

The optimal choice for Ω is $\Omega \propto V_0^{-1}$, and with such an Ω, $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, (\Gamma_0' V_0^{-1} \Gamma_0)^{-1})$.

A standard two-step GMM procedure can be applied to obtain an HIV estimator based on the optimal GMM weighting matrix, where $\Omega_n \equiv \left[\frac{1}{n}\sum_{i=1}^n \psi(\hat{\epsilon}_{i,n})^2 q_{i,n} q_{i,n}'\right]^{-1}$ for the optimal GMM estimation, where $\hat{\epsilon}_{i,n} \equiv y_{i,n} - z_{i,n}'\tilde{\theta}$ and $\tilde{\theta}$ is an initial consistent estimator. We can estimate Γ_0 and V_0 by their sample means $\frac{1}{n}\sum_{i=1}^n \dot{\psi}(\hat{\epsilon}_{i,n}) q_{i,n} z_{i,n}'$ and $\frac{1}{n}\sum_{i=1}^n \psi(\hat{\epsilon}_{i,n})^2 q_{i,n} q_{i,n}'$, respectively (see Theorem 5 for their consistency).

3.3. Large Sample Properties of the Huber GMM Estimator

For the HGMM estimator, we require the linear Huber moments to satisfy the conditions for the HIV estimators. In addition, some conditions for the quadratic Huber moments are also needed. We use W_n, $W_n^2 - \mathrm{diag}(W_n^2)$, $S_n^{-1} - \mathrm{diag}(S_n^{-1})$, or

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

$G_n - \text{diag}(G_n)$ as P_{kn} in this article. Since

$$\psi(\epsilon_n(\theta))' P_{kn} \psi(\epsilon_n(\theta)) = \sum_{i=1}^{n} \psi(\epsilon_{i,n}(\theta)) \left[P_{i\cdot,kn} \psi(\epsilon_n(\theta)) \right], \tag{3.1}$$

if $\left\{ P_{i\cdot,kn} \psi(\epsilon_n(\theta)) \right\}_{i \in D_n}$ is uniformly bounded and uniformly L_2-NED on $\{x_{i,n}\epsilon_{i,n} : i \in D_n\}$, by Lemma A.5, the product term $\left\{ \psi(\epsilon_{i,n}(\theta)) \left[P_{i\cdot,kn} \psi(\epsilon_n(\theta)) \right] \right\}_{i \in D_n}$ is also uniformly bounded and uniformly L_2-NED. Then, we can investigate the limiting properties of the quadratic Huber moments using the NED theory. We show that $\left\{ P_{i\cdot,kn} \psi(\epsilon_n(\theta)) \right\}_{i \in D_n}$ is uniformly L_2-NED on $\{x_{i,n}\epsilon_{i,n} : i \in D_n\}$ in Assumption A.9. Thus, we are ready to establish asymptotic properties of the HGMM estimator.

THEOREM 3. *Under Assumptions 1–8,* $\hat{\theta}_{HGMM} \xrightarrow{p} \theta_0$.

Next, we obtain the asymptotic distribution of the HGMM estimator $\hat{\theta}_{HGMM}$. To do so, we need some regularity conditions for $\Gamma_n(\theta) \equiv \frac{\partial g_n(\theta)}{\partial \theta'}$, $\Gamma_n \equiv \Gamma_n(\theta_0)$, and $V_n \equiv n \,\text{var}\,(g_n(\theta_0))$. We collect the expressions of these matrices in Section C of the Supplementary Material. Notice that the V_n in Eq. (C.2) is a block diagonal matrix. This is because a linear form and a quadratic form (with all diagonal elements being zero) are uncorrelated, as suggested by Eq. (3.2) in Kelejian and Prucha (2001).

Assumption 12. The same conditions as in Assumption 10 hold for the $g_n(\theta)$ in Eq. (2.6).

Assumption 13. $\lim_{n\to\infty} V_n \equiv \lim_{n\to\infty} n \,\text{var}\,(g_n(\theta_0)) = V_0$ for some positive-definite matrix V_0.

THEOREM 4. *Under Assumptions 1–9, 12, and 13, if the* α *in Assumption 4(ii) satisfies* $\alpha > \left[\frac{2}{\min(2,\gamma)} + 1 \right] d$, *then*

$$\sqrt{n}(\hat{\theta}_{HGMM} - \theta_0) \xrightarrow{d} N(0, (\Gamma_0' \Omega \Gamma_0)^{-1} \Gamma_0' \Omega V_0 \Omega \Gamma_0 (\Gamma_0' \Omega \Gamma_0)^{-1}).$$

When $\Omega \propto V_0^{-1}$, $\sqrt{n}(\hat{\theta}_{HGMM} - \theta_0) \xrightarrow{d} N(0, (\Gamma_0' V_0^{-1} \Gamma_0)^{-1})$.

The condition $\alpha > \left[\frac{2}{\min(2,\gamma)} + 1 \right] d$ ensures that the NED coefficient satisfies the condition of Lemma A.3 in Jenish and Prucha (2012). If $|w_{ij,n}|$ decreases exponentially as d_{ij} increases, this condition becomes unnecessary.

Following Lin and Lee (2010), we estimate V_0 in Eq. (C.2) in the Supplementary Material by replacing $\Sigma_n \equiv \text{var}(\psi(\epsilon_n))$ by $\hat{\Sigma}_n \equiv \text{diag}(\psi(\hat{\epsilon}_{1,n})^2, \ldots, \psi(\hat{\epsilon}_{n,n})^2)$ and replacing Σ_n^d by $\hat{\Sigma}_n^d \equiv (\psi(\hat{\epsilon}_{1,n})^2, \ldots, \psi(\hat{\epsilon}_{n,n})^2)'$, where $\hat{\epsilon}_{i,n} = y_{i,n} - z_{i,n}' \hat{\theta}$ is the

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

residual, i.e., the estimator for V_0 is

$$\hat{V} = \frac{1}{n}\begin{pmatrix} (\hat{\Sigma}_n^d)'(P_{1n} \odot P_{1n}^s)\hat{\Sigma}_n^d & \cdots & (\hat{\Sigma}_n^d)'(P_{1n} \odot P_{mn}^s)\hat{\Sigma}_n^d & 0 \\ \vdots & \vdots & \vdots & \vdots \\ (\hat{\Sigma}_n^d)'(P_{mn} \odot P_{1n}^s)\hat{\Sigma}_n^d & \cdots & (\hat{\Sigma}_n^d)'(P_{mn} \odot P_{mn}^s)\hat{\Sigma}_n^d & 0 \\ 0 & \cdots & 0 & Q_n'\hat{\Sigma}_n Q_n \end{pmatrix}.$$

Similarly, we estimate Γ_n in Eq. (C.1) in the Supplementary Material by

$$\hat{\Gamma}_n = -\frac{1}{n}\begin{pmatrix} \dot{\psi}(\hat{\epsilon}_n)'(P_{1n}^s \odot G_n)[\psi(\hat{\epsilon}_n) \odot \hat{\epsilon}_n] & 0 \\ \vdots & \vdots \\ \dot{\psi}(\hat{\epsilon}_n)'(P_{mn}^s \odot G_n)[\psi(\hat{\epsilon}_n) \odot \hat{\epsilon}_n] & 0 \\ [(\dot{\psi}(\hat{\epsilon}_n)\iota_q') \odot Q_n]'W_n Y_n & [(\dot{\psi}(\hat{\epsilon}_n)\iota_q') \odot Q_n]'X_n \end{pmatrix}.$$

Then, we can estimate the asymptotic variance of the optimal GMM estimator by $(\hat{\Gamma}_n'\hat{V}^{-1}\hat{\Gamma}_n)^{-1}$.

THEOREM 5. *Suppose Assumptions 1–9, 12, and 13 hold. And conditional on $x_{i,n}$, the probability density function (p.d.f.) of $\epsilon_{i,n}$, denoted by $f_{in}(\cdot|x_{i,n})$, satisfies $\sup_{i,n,\epsilon,x}|f_{in}(\epsilon|x)| < \infty$ and $\sup_{i,n,\epsilon,x}|\epsilon f_{in}(\epsilon|x)| < \infty$. Suppose $\gamma \geq 2$. Then, (1) $\hat{V} \xrightarrow{p} V_0$ and (2) $\hat{\Gamma}_n \xrightarrow{p} \Gamma_0$.*

The conditions $\sup_{i,n,\epsilon,x}|f_{in}(\epsilon|x)| < \infty$ and $\sup_{i,n,\epsilon,x}|\epsilon f_{in}(\epsilon|x)| < \infty$ are satisfied by most continuous and piecewise continuous p.d.f.s in practice. As long as $f_{in}(x)$ is continuous and $\lim_{|x|\to\infty}|xf_{in}(x)| = 0$, Assumption 2(ii) holds. Even when $f_{in}(\cdot)$ is not continuous, Assumption 2(ii) is still satisfied in many cases, e.g., the p.d.f. of an exponential distribution or a uniform distribution.

3.4. Finite Sample Breakdown Point Analysis

Breakdown point analysis is a key criterion for assessing the robustness of a statistical procedure. "Informally, a breakdown of a statistical procedure means that the procedure no longer conveys useful information on the data-generating mechanism" (Genton and Lucas, 2003). The concept of breakdown point varies across different literature. Traditionally, the breakdown point refers to the fraction of contamination (or outliers) that results in unbounded estimators or estimators on the boundary of the parameter space (Hampel, 1971). This definition is well-suited for independent data, but Genton and Lucas (2003) argue that it is not readily applicable to dependent data. Thus, they extend the definition in Hampel (1971) to accommodate dependent data. Genton, (2003, p. 151) finds that when the spatial weights matrix is lower triangular, the breakdown point for the MLE of a pure SAR model without exogenous variables is zero; when $w_{ij} = 1$ if and only if $|i-j| = 1$, the asymptotic breakdown point of the least median of squares estimator for the pure SAR model with additive outliers is 14.4%.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

Since "the breakdown point is most useful in small sample situations" (Huber and Ronchetti, 2009, p.279), we focus on finite sample breakdown point. As in Genton and Lucas (2003), we consider innovation outliers, i.e., outliers in the error terms $\epsilon_{i,n}$.[12] Unlike additive outliers used in Genton (2003), the impact of innovation outliers can be propagated through the spatial weights matrix W_n to all dependent variables $y_{j,n}$'s. Assuming the presence of contamination $\zeta\bar{\epsilon}_n$, where $\zeta \in [0,\infty]$ controls the magnitude of the contamination, the model then becomes

$$Y_n = \lambda_0 W_n Y_n + X_n\beta_0 + (\epsilon_n + \zeta\bar{\epsilon}_n).$$

So, $Y_n = S_n^{-1}(X_n\beta + \epsilon_n + \zeta\bar{\epsilon}_n)$. Suppose all entries of W_n are nonnegative and $\lambda_0 > 0$ to simplify our analysis. We consider two cases:

(i) The network implied by W_n is bidirectional and connected, i.e., W_n cannot be written as a block diagonal matrix. In this case, for any $i \neq j$, there exists a $k \in \mathbb{N}$ such that $\left(W_n^k\right)_{ij} \neq 0$. Consequently, all entries of $S_n^{-1} = \sum_{l=0}^{\infty}\lambda_0^l W_n^l$ are greater than 0. Therefore, even if there exists only one outlier, e.g., $\bar{\epsilon}_n = (1,0,\ldots,0)'$, as $\zeta \to +\infty$, all $y_{i,n}$'s will tend to $+\infty$. In this case, it is impossible to obtain any reasonable estimators. Thus, the breakdown point for innovations outliers is zero. This outcome is independent of any specific estimation method and is attributed to the network's connectedness. This conclusion parallels the zero breakdown point for the SAR model discussed in Genton (2003, p. 151).

(ii) W_n is a block diagonal matrix. Suppose the network can be separated into $G > 1$ subnetworks, represented by W_g $(g = 1,\ldots,G)$. These subnetworks are not linked, but each of them is a bidirectional and connected network, i.e., $W_n = \mathrm{diag}(W_1,\ldots,W_G)$. In each subnetwork W_g $(g = 1,\ldots,G)$, if we introduce one extreme outlier to one $\epsilon_{i,n}$, where i is an individual in group g, i.e., the contaminated error is $\epsilon_{i,n} + \zeta$. Since W_g is connected, by an argument similar to case (i), as $\zeta \to +\infty$, all $y_{j,n}$'s in group g will tend to $+\infty$. Therefore, the finite sample breakdown point must be $\leq \frac{G-1}{n}$. Consequently, if the number of groups G is not large, the finite sample breakdown point is quite small. When G is large, the data are more similar to independent data, and it is reasonable to have a greater breakdown point.

In summary, due to network structures, when extreme contamination occurs in innovations, it will also lead to extreme values for other observations. As a result, the breakdown point is smaller than that in the typical case of independent data.

3.5. Influence Function

The influence function is another useful criterion for measuring the robustness of an estimator. In this section, we will demonstrate that, under regularity conditions, the influence functions of both the HIV estimator and the HGMM estimator are bounded.

[12] Similar analysis is also applicable to outliers in $x_{i,n}$.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

Consider the influence functions of the HIV estimator first. Suppose that with probability $1-\delta$, (Y_n, X_n) is generated by an uncontaminated model, and with probability δ, $(Y_n, X_n) = (Y_n^0, X_n^0)$, where (Y_n^0, X_n^0) is a nonstochastic outlier. With such contaminated data, the pseudo true value of the parameter θ is defined as

$$\theta_\delta = \arg\min \left\{ \frac{1-\delta}{n} \sum_{i=1}^{n} \mathrm{E}\left[q_{i,n} \psi(y_{i,n} - z_{i,n}'\theta) \right] + \frac{\delta}{n} \sum_{i=1}^{n} \left(\mathrm{E} q_{i,n} \right) \psi(y_{i,n}^0 - (z_{i,n}^0)'\theta) \right\}' \Omega_n \left\{ \frac{1-\delta}{n} \sum_{i=1}^{n} \mathrm{E}\left[q_{i,n} \psi(y_{i,n} - z_{i,n}'\theta) \right] + \frac{\delta}{n} \sum_{i=1}^{n} \left(\mathrm{E} q_{i,n} \right) \psi(y_{i,n}^0 - (z_{i,n}^0)'\theta) \right\}, \tag{3.2}$$

where the expectation E is taken for the uncontaminated data and $z_{i,n}^0 \equiv (w_{i\cdot,n} Y_n^0, (x_{i,n}^0)')'$. The influence function is defined as

$$\lim_{\delta \to 0} \frac{\theta_\delta - \theta_0}{\delta}.$$

Assume that $\lim_{\delta \to 0} \theta_\delta = \theta_0$, which is a standard condition for calculating influence functions, and that $\psi(y_{i,n}^0 - (z_{i,n}^0)'\theta)$ is differentiable around θ_0. The first-order condition of Eq. (3.2) is

$$\left\{ \frac{1-\delta}{n} \sum_{i=1}^{n} \mathrm{E}\left[\dot{\psi}(y_{i,n} - z_{i,n}\theta_\delta) q_{i,n} z_{i,n}' \right] - \frac{\delta}{n} \sum_{i=1}^{n} \mathrm{E}\left[q_{i,n} (z_{i,n}^0)' \right] \dot{\psi}(y_{i,n}^0 - (z_{i,n}^0)'\theta_\delta) \right\}' \cdot \Omega_n \left\{ \frac{1-\delta}{n} \sum_{i=1}^{n} \mathrm{E}\left[q_{i,n} \psi(y_{i,n} - z_{i,n}\theta_\delta) \right] + \frac{\delta}{n} \sum_{i=1}^{n} \left(\mathrm{E} q_{i,n} \right) \psi(y_{i,n}^0 - (z_{i,n}^0)'\theta_\delta) \right\} = 0.$$

Doing Taylor expansion for the second term at θ_0 in the above equation, we obtain

$$\left\{ \frac{1-\delta}{n} \sum_{i=1}^{n} \mathrm{E}\left[\dot{\psi}(y_{i,n} - z_{i,n}\theta_\delta) q_{i,n} z_{i,n}' \right] - \frac{\delta}{n} \sum_{i=1}^{n} \mathrm{E}\left[q_{i,n} (z_{i,n}^0)' \right] \psi(y_{i,n}^0 - (z_{i,n}^0)'\theta_\delta) \right\}' \Omega_n \cdot \left\{ \frac{1-\delta}{n} \sum_{i=1}^{n} \left(\mathrm{E}\left[\psi(\epsilon_{i,n}) q_{i,n} z_{i,n}' \right] + \mathrm{E}\left[\dot{\psi}(y_{i,n} - z_{i,n}\bar{\theta}) q_{i,n} z_{i,n}' \right] (\theta_0 - \theta_\delta) \right) + \frac{\delta}{n} \sum_{i=1}^{n} \left(\mathrm{E} q_{i,n} \right) \psi(y_{i,n}^0 - (z_{i,n}^0)'\theta_\delta) \right\} = 0,$$

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

where $\bar{\theta}$ is between θ_0 and θ_δ.[13] Hence, as $\delta \to 0$,

$$\frac{\theta_\delta - \theta_0}{\delta} \to \left\{ \left(\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[\dot{\psi}(\epsilon_{i,n}) z_{i,n} q_{i,n}'\right] \right) \Omega_n \left(\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[\dot{\psi}(\epsilon_{i,n}) q_{i,n} z_{i,n}'\right] \right) \right\}^{-1} \cdot \left(\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[\dot{\psi}(\epsilon_{i,n}) z_{i,n} q_{i,n}'\right] \right) \Omega_n \left[\frac{1}{n}\sum_{i=1}^{n} \left(\mathrm{E}\, q_{i,n}\right) \psi\left(y_{i,n}^0 - (z_{i,n}^0)'\theta_0\right) \right].$$

It follows from $|\psi(\cdot)| \le M$ that $\frac{\theta_\delta - \theta_0}{\delta}$ is bounded as $\delta \to 0$. In addition, under Assumptions 7 and 10, because $\sup_{i,n} \mathrm{E}\left\|q_{i,n}\right\| < \infty$, we have $\limsup_{n\to\infty} \mathrm{E} \lim_{\delta\to 0} \left\| \frac{\theta_\delta - \theta_0}{\delta} \right\| < \infty$. In other words, the bound for the influence function of the HIV estimator for each n remains bounded as $n \to \infty$.

By similar arguments, the influence function of the HGMM estimator is also bounded.

4. MONTE CARLO SIMULATION

In this section, we assess finite sample performances of our proposed estimators using Monte Carlo simulations. In most experiments, we consider the following SAR model: $y_{i,n} = \lambda_0 w_{i\cdot,n} Y_n + \beta_{10} + \beta_{20} x_{i2,n} + \beta_{30} x_{i3,n} + \epsilon_{i,n}$, where $x_{i2,n} \overset{iid}{\sim} N(0,1)$, $(x_{13,n}, \ldots, x_{n3,n})' \sim (I_n - 0.2W_n)^{-1} N(0, I_n)$, $\lambda_0 = 0.5$, and $\beta_{10} = \beta_{20} = \beta_{30} = 1$. Notice that we permit spatial dependence among $x_{i3,n}$'s. The spatial weights matrix W_n is constructed from some counties of the United States, and it is row-normalized.[14] Additional simulation studies can be found in Section E of the Supplementary Material.

4.1. The Effects of Different *c* in *M*

In Section 2.3, we propose to use Eq. (2.7) to determine the critical value M: $M = c \cdot \frac{\text{median}(|\epsilon - \text{median}(\epsilon)|)}{0.6745}$, where $c = 1.345$, the default critical value in both R and MATLAB. We first conduct some simulations to examine the performance of the HIV estimator and the HGMM estimator under different choices of c.

We generate independent disturbances $\epsilon_{i,n}$'s from the following seven distributions: (1) i.i.d. $N(0,1)$; (2) a mixed normal distribution $0.95N(0,1) + 0.05N(0,25)$; (3) a mixed distribution of $N(0,1)$ and the standard Cauchy distribution $0.95N(0,1) + 0.05t_1$; (4) $N(0,1)$ contaminated by a single extreme error term: $\epsilon_i \sim N(0,1)$, $i = 2,3,\ldots,n$ and $\epsilon_1 \sim N(150,1)$; (5) $N(0,1)$ contaminated by two extreme error terms: $\epsilon_i \sim N(0,1)$, $i = 3,4,\ldots,n$ and $\epsilon_1, \epsilon_2 \sim N(150,1)$; (6) Laplace distribution times $x_{i2,n}$; and (7) exp(1) times $x_{i2,n}^2$. Setups (2)–(7) are contaminated or long-tailed distributions. Setups (2) and (3) belong to mixture

[13] For different rows of this equation, the values of $\bar{\theta}$ might be different.

[14] See Xu and Lee (2015a) for details.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

distributions, and setups (4) and (5) are contaminated by extreme outliers.[15] Setups (6) and (7) are long-tailed and subject to conditional heteroskedasticity. Estimation results are summarized in Tables E.1–E.3 in the Supplementary Material.

For the HIV estimator, we consider a single tuning parameter c under the several different types of disturbances with a sample size of $n = 100$. As the default value of c is 1.345 in R and MATLAB, we take this value as a benchmark to calculate the relative mean squared error (RelaMSE) for each tuning parameter value, i.e., $RelaMSE = \frac{MSE(\hat{\theta}_c)}{MSE(\hat{\theta}_{1.345})}$. The results are shown in Table E.1 in the Supplementary Material. In most cases, for contaminated or long-tailed distributions, the HIV estimator performs better when $c \in [1, 1.5]$ than when $c > 1.5$. Within the range $[1, 1.5]$, the relative efficiency varies by less than 10% across different tuning parameters, with the exception of the case involving Cauchy contamination. As a comparison, from Table 3 in Section 4.2, the relative efficiency gain of the HIV estimator over 2SLS is at least 86% (in most cases more than 100%) when $n = 100$. Therefore, the efficiency difference due to varying c within [1, 1.5] is minor, compared to the efficiency gained by using Huber estimation. Notably, $c = 1.345$ works substantially better than other values of c in the presence of Cauchy contamination. Hence, we use $c = 1.345$ for all Huber IV estimators in subsequent experiments.

For the HGMM estimator, we adopt two tuning parameters: c_1 for quadratic moments and c_2 for linear moments, where $c_1, c_2 \in \{1, 1.2, 1.345, 1.5, 1.8, 2, 2.5, 3\}$. We explore the impact of using different pairs of tuning parameters in the following three scenarios: (1) disturbances drawn from $0.95N(0, 1) + 0.05N(0, 25)$, (2) $N(0, 1)$ contaminated by a single outlier from $N(150, 1)$, and (3) Laplace distribution times $x_{i2,n}$. Similarly, we use the pair (1.345, 1.345) as a benchmark to calculate the relative MSE for each tuning parameter combination, i.e., $RelaMSE = \frac{MSE(\hat{\theta}_{(c_1,c_2)})}{MSE(\hat{\theta}_{(1.345,1.345)})}$. The results regarding λ and β_2 are presented in Tables E.2 and E.3 in the Supplementary Material. In general, the relative MSEs of both $\hat{\lambda}$ and $\hat{\beta}_2$ tend to get larger when $c_2 \geq 1.8$ and do not vary much for $c_2 \in [1, 1.5]$. For a fixed c_2, especially when $c_2 \in [1, 1.5]$, the influence of c_1 on the relative MSE is rather small, with variations typically within 5% in most cases. In cases where there is no clear pattern, such as the relative RMSE of $\hat{\lambda}$ in scenario (2), $(c_1, c_2) = (1.345, 1.345)$ dominates most other tuning parameter combinations. Thus, Eq. (2.7) serves as a reasonable criterion for selecting M with $(c_1, c_2) = (1.345, 1.345)$, and there is no need to differentiate c_1 and c_2.

In conclusion, for both the HIV estimator and the HGMM estimator, our simulation results indicate that $c = 1.345$ is a satisfactory choice for a variety of scenarios, including contaminated data, long-tailed data, and data with conditional heteroskedasticity. Thus, we maintain $c = 1.345$ in our subsequent simulation studies.

[15] We borrow the extreme contamination setup from Tho et al. (2023). We thank one of the anonymous referees for the suggestion.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

4.2. Compare the Performance of Various Estimators

Next, we compare the performances of the QMLE, the 2SLS estimator in Kelejian and Prucha (1998), the BIV estimator in Lee (2003), the GMM estimators in Lin and Lee (2010), the quantile IV (QIV) estimator in Su and Yang (2011), the 2SGM estimator in Wagenvoort and Waldmann (2002),[16] the Huber maximum likelihood estimator (HMLE) in Tho et al. (2023), and the three estimators proposed in our article: the HIV estimator, the BHIV estimator, and HGMM estimator. For both the GMM estimator and the HGMM estimator, we adopt $P_n = \tilde{G}_n - \text{diag}(\tilde{G}_n)$ for the quadratic moments, where $\tilde{G}_n = W_n(I_n - \tilde{\lambda}W_n)^{-1}$, and $\tilde{\lambda}$ is a consistent estimator for λ_0.[17] And we use X_n and $G_nX_n\tilde{\beta}$ as the IVs, where $\tilde{\beta}$ is a consistent estimator for the true parameter value β_0. For Huber type estimators, the critical value of M is determined as in Section 2.3 with $c = 1.345$.

We have tried three different sample sizes in the experiments: $n = 100, 200,$ and 400. For each setting, simulations are repeated 1,000 times to obtain these estimates' sample bias and RMSE. We report the results on parameters λ and β_2 under sample sizes $n = 100$ and $n = 400$ in Tables 1 and 2, respectively. Full results—including all parameters (except for the intercept coefficient β_1) and all sample sizes ($n = 100, 200, 400$)—are provided in Tables E.4–E.6 in the Supplementary Material. Under asymmetrically distributed disturbances, the true intercept β_{10} depends on the critical value M, and is different from the true intercept under the moment condition $\text{E}\,\epsilon_{i,n} = 0$, so it is troublesome to compare its estimates. Thus, we omit them from all reported tables—including the full results in Tables E.4–E.6 in the Supplementary Material. To obtain a better insight into the merits of robust estimators, we present the relative efficiency of each Huber-type estimator $\hat{\theta}$ compared to its corresponding traditional estimator, e.g., HIV/2SLS, BHIV/BIV, and HGMM/GMM, in Table 3. The relative efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is calculated as $MSE(\hat{\theta}_2)/MSE(\hat{\theta}_1)$.

Below are some interesting observations obtained from the simulation results:

(1) For all sample sizes and distributions, the biases and RMSEs of the three proposed Huber estimators are notably small. As the sample size n increases, the RMSEs of our estimators decrease at roughly the rate of $n^{-1/2}$. For instance, under the mixed normal distribution, when $n = 100$, 200, and 400, the RMSE for $\hat{\lambda}_{HGMM}$ is 0.0661, 0.0473, and 0.0378, respectively. These results conform with the prediction of our large sample theories.

(2) When the disturbances are i.i.d. normally distributed,

$$QMLE \succsim GMM \succsim HGMM \succsim HMLE \succ 2SLS \approx BIV \approx HIV \approx BHIV \succ 2SGM \succ QIV,$$

[16] The 2SGM estimation of the linear regression model $Y_n = X_n\beta_0 + \epsilon_n$ with endogenous variables X_n can be done in the following procedure: (1) Run a robust estimation of each column of the endogenous variables X_{jn} on IVs Q_n to obtain an estimate of X_{jn}: $\hat{X}_{jn} = Q_n[Q_n'W(\hat{e}_j)Q_n]^{-1}Q_n'W(\hat{e})X_{jn}$, where $W(\hat{e}_j)$ is a diagonal matrix giving outliers among $\hat{e}_{ji}$'s lower weights, where $\hat{e}_{ji}$'s are the residuals of regressing X_j on Q_n and (2) run a similar robust estimation of Y_n on $\hat{X}_n$. Wagenvoort and Waldmann (2002) construct a weight function using Tukey's bi-square function.

[17] For the GMM estimator, we use the 2SLS estimate for $\tilde{\lambda}$ and $\tilde{\beta}$; for the HGMM estimator, we use the HIV estimate for them.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

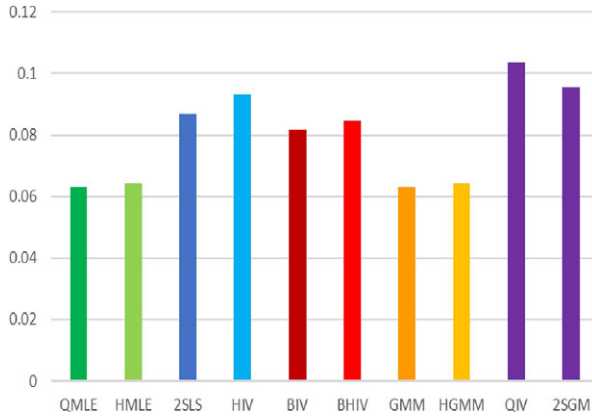


FIGURE 1. RMSE of $\hat{\lambda}$ of various estimators under $N(0,1)$.

where $\succ$ means "performing better than," $\succsim$ means "performing slightly better than," and $\approx$ means "performing similarly." Although QMLE performs the best, the differences between the HGMM estimator, the GMM estimator, and the QMLE are minor. Among all robust estimators, the HGMM estimator exhibits the smallest efficiency loss under $N(0,1)$. For all sample sizes and all parameters,

$$\frac{RMSE(HGMM) - RMSE(QMLE)}{RMSE(QMLE)} < 2\%,$$

$$\frac{RMSE(HGMM) - RMSE(GMM)}{RMSE(GMM)} < 2\%.$$

From Table 3, under normally distributed disturbances, Huber-type estimators exhibit only a slight loss in efficiency (in most cases $< 6\%$) compared to traditional estimators. This modest efficiency trade-off is to be expected, as robustness often comes at the cost of some efficiency. The efficiency losses incurred by our estimators are minimal. The aforementioned observations can be clearly seen from Figures 1 and 2. Please note that each pair of bars sharing similar colors in the figure represents a traditional estimator and its Huber variant, respectively (for instance, 2SLS and HIV, BIV and BHIV, and GMM and HGMM estimators), while the purple bars represent 2SGM estimator and QIV estimator.

(3) In extreme contamination scenarios (cases 4 and 5), the biases and RMSEs of the traditional estimators, including QMLE, 2SLS, BIV, and GMM, are extremely large, among which BIV is the worst (when $n = 100$, for λ, the efficiency of the BHIV estimator compared to the BIV estimator is 3.2×10^5). In contrast, all robust estimators exhibit significantly superior performance. The Huber MLE is slightly better than HGMM, but the difference is minimal. Both significantly outperform other robust estimators. Overall, we summarize performances of these estimators

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

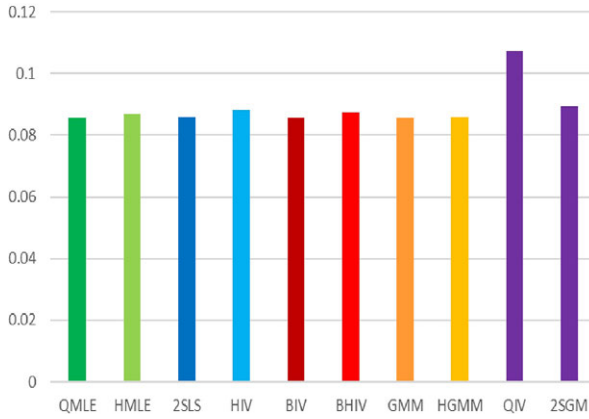


FIGURE 2. RMSE of $\hat{\beta}_2$ of various estimators under $N(0,1)$.

in most (although not all) cases as follows:

$$HMLE \succsim HGMM \succ BHIV \succ HIV \succ QIV \succ 2SGM \succ QMLE \succ 2SLS \approx GMM \succ BIV.$$

The conclusions for setups (2) and (3) are similar.

(4) In the presence of conditional heteroskedasticity (cases 6 and 7), the HGMM estimator and QIV estimator outperform other robust estimators, which in turn perform better than traditional estimators. We can see this more clearly from Figures 3 and 4. Results can be summarized as follows:

$$HGMM \approx QIV \succ BHIV \succ HIV \approx HMLE \succ 2SGM \succ GMM \approx QMLE \succ 2SLS \succ BIV.$$

(5) In all setups (2)–(7), which involve contaminated, long-tailed, or conditionally heteroskedastic disturbances, Huber-type estimators consistently outperform their traditional counterparts:

$$\begin{cases} HGMM \succ GMM, \\ BHIV \succ BIV, \\ HIV \succ 2SLS. \end{cases}$$

(6) All the proposed estimators are computationally efficient. For example, when the errors follow a mixed normal distribution and the sample size $n = 100$, our desktop computer (CPU: Intel(R) Xeon(R) E-2136; RAM: 32G; hard drive: 1T; and software: MATLAB) takes an average of 0.013 seconds to run the HGMM estimation and 0.25 seconds for the Huber MLE by Tho et al. (2023). With $n = 400$, the HGMM estimation takes 0.047 seconds, while the Huber MLE takes 5 seconds.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

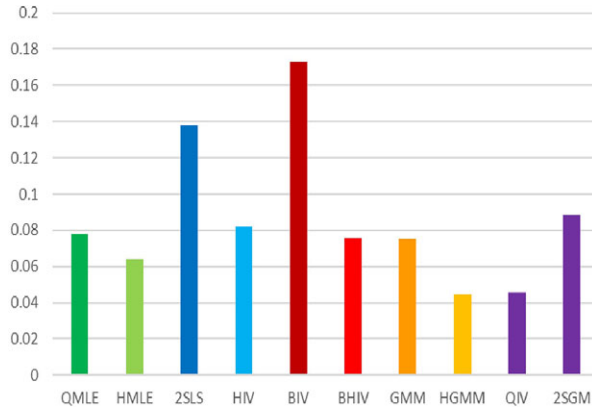


FIGURE 3. RMSE of $\hat{\lambda}$ of various estimators under $Laplace \cdot x_{i2}$.

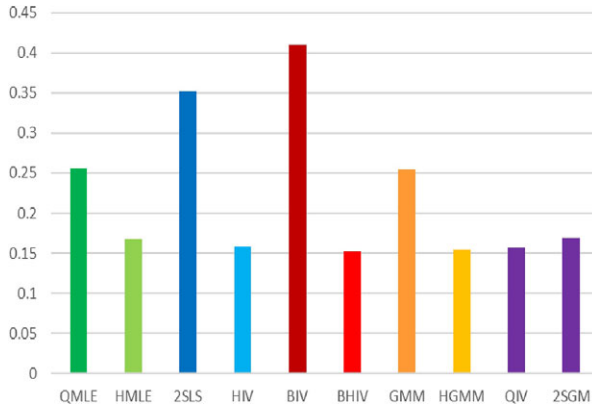


FIGURE 4. RMSE of $\hat{\beta}_2$ of various estimators under $Laplace \cdot x_{i2}$.

On average, our estimators can be computed 20 times faster than the estimator in Tho et al. (2023) for $n = 100$, and 100 times faster for $n = 400$.

In conclusion, the simulation outcomes demonstrate that our estimators are robust against both contaminated or long-tailed disturbances and conditional heteroskedasticity, incur minimal efficiency loss under short-tailed disturbances, and are computationally convenient. The Huber MLE by Tho et al. (2023) works well under contaminated disturbances, but is less robust to conditional heteroskedasticity, and computationally slower. QIV estimator excels under heteroskedastic long-tailed disturbances, but performs worse than the HGMM estimator and the Huber MLE in other scenarios. Overall, HGMM estimator performs the best among all these estimators in simulations.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

4.3. A Test Experiment

We further carry out simulation studies to evaluate the empirical size and the power of testing $H_0 : \beta_{20} = \beta_{30}$ vs. $H_0 : \beta_{20} \neq \beta_{30}$ at level 5%, as suggested by an anonymous referee. The experimental setup closely follows that of Section 4.3, except that $\beta_{30} \in \{1, -1\}$. Detailed results are reported in Tables E.7 and E.8 in the Supplementary Material. We observe that the empirical size is around 0.05 and the empirical power is close to 1 for all settings and all sample sizes. As the sample size increases, the empirical size gets closer to 0.05 and the empirical power tends toward 1. This shows that our asymptotic distribution approximates the true distribution of the estimators well.

5. EFFECT OF URBAN HEAT ISLAND ON HOUSING PRICES

From the Monte Carlo study, it is evident that the HGMM estimation significantly outperforms in scenarios characterized by long-tailed disturbances and conditional heteroskedasticity. In this section, we employ the HGMM approach to identify the effect of urban heat island on surrounding housing prices in Beijing of China, using the data in 2017 from Lin et al. (2022). In our analysis, each individual is a community, the explained variable is the logarithm of the average housing prices in a community, and the variable of interest is an index of urban heat island intensity (UHII).

The urban heat island effect refers to a phenomenon in which the temperature in urban areas is higher than that in the surrounding countryside (Mohajerani, Bakaric, and Jeffrey-Bailey, 2017). Due to the high temperature near the ground in the central area of the heat island, the atmosphere moves upward, forming a pressure difference with the surrounding areas. The near-ground atmosphere in the surrounding areas converges to the central area, forming a low-pressure vortex in the central area of the city, resulting in the accumulation of sulfur oxides, nitrogen oxides, carbon oxides, hydrocarbons, and other air pollutants formed by burning fossil fuels in people's lives, industrial production, and vehicle operation in the central area of the heat island (Zhong et al., 2017). The warming effect of urbanization is believed to pose a major risk to public health, such as emotional irritability, mental depression, insomnia, anorexia, dyspepsia, increased ulcers, recurrence of gastrointestinal diseases, etc. (Hsu et al., 2021). The urban heat island effect in high-temperature weather also increases the possibility for residents to use air conditioners, electric fans, and other electrical appliances, thus bolstering electricity consumption, increasing the power load, and generating more waste heat to further enhance the heat island effect. Due to its critical impact on human health and comfort, the urban heat island effect can be capitalized into housing prices by affecting residents' choices in home purchasing.

Plantinga et al. (2013) find that residents prefer higher January temperatures but dislike higher July temperatures. Klaiber, Abbott, and Smith (2017) hold that homeowners value cooler temperatures. Lin et al. (2022) establish an index for

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

TABLE 1. Estimation results ($n = 100$)

	Dist.	1: $N(0,1)$	2: Mixed normal	3: NC	4: Extreme I	5: Extreme II	6: Laplace$\cdot x_{i2}$	7: exp(1)$\cdot x_{i2}^2$
					λ			
QMLE	Bias	−0.0012	−0.0210	−0.201	−0.0838	−0.1105	−0.0197	−0.0243
	RMSE	0.0630	0.0784	0.0828	0.0845	0.1108	0.0776	0.0221
2SLS	Bias	−0.0012	0.0049	0.0293	0.6074	0.6347	0.0133	0.0478
	RMSE	0.0867	0.1333	0.2517	0.6239	0.6404	0.1380	0.1949
BIV	Bias	−0.0061	−0.0096	−0.0588	11.1149	0.4889	−0.0174	−0.0407
	RMSE	0.0818	0.1317	2.0752	34.2853	3.1589	0.1730	0.4714
GMM	Bias	−0.0099	−0.0134	0.0121	0.3805	0.5507	−0.0134	−0.0152
	RMSE	0.0632	0.0766	0.4555	0.6415	0.6604	0.0750	0.0318
QIV	Bias	−0.0012	0.0049	0.0069	0.0045	0.0114	0.0044	0.0032
	RMSE	0.1036	0.0971	0.0966	0.0700	0.0522	0.0455	0.0421
2SGM	Bias	−0.0015	0.0017	−0.0035	0.0308	0.1791	−0.0056	−0.0034
	RMSE	0.0955	0.1119	0.3233	0.0867	0.2313	0.0882	0.0911
HMLE	Bias	−0.0148	−0.0178	−0.0099	0.0007	−0.0065	−0.0273	−0.0145
	RMSE	0.0643	0.0648	0.0626	0.0425	0.0571	0.0643	0.0456
HIV	Bias	0.0005	−0.0051	−0.0006	0.0046	0.0430	−0.0050	0.0012
	RMSE	0.0933	0.1038	0.0987	0.0607	0.1329	0.0819	0.0717
BHIV	Bias	0.0066	−0.0060	−0.0025	−0.0038	0.0378	−0.0055	0.0035
	RMSE	0.0846	0.0912	0.0904	0.0606	0.1357	0.0757	0.0536
HGMM	Bias	−0.0105	−0.0131	−0.0059	0.0105	0.0365	−0.0091	−0.0059
	RMSE	0.0642	0.0661	0.0647	0.0496	0.0558	0.0493	0.0318

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

TABLE 1. Continued

	Dist.	1: $N(0,1)$	2: Mixed normal	3: NC	4: Extreme I	5: Extreme II	6: Laplace$\cdot x_{i2}$	7: exp(1)$\cdot x_{i2}^2$
					β_2			
QMLE	Bias	0.0025	−0.0017	−0.1610	1.9731	−0.7563	0.0027	−0.0076
	RMSE	0.0855	0.1470	4.3334	1.9750	0.7613	0.2560	0.5098
2SLS	Bias	0.0004	−0.0138	−0.3094	2.0925	−0.3651	−0.0040	0.0871
	RMSE	0.0857	0.1501	8.9806	2.0943	0.3821	0.3524	0.4955
BIV	Bias	0.0010	−0.0107	−0.2055	5.1444	−0.4518	−0.0036	0.1029
	RMSE	0.0856	0.1507	7.1583	9.8163	1.9376	0.4098	0.5374
GMM	Bias	−0.0003	−0.0090	−0.2808	2.0615	−0.4518	−0.0008	0.1052
	RMSE	0.0856	0.1480	7.7919	2.0656	0.4739	0.2550	0.5091
QIV	Bias	−0.0007	−0.0008	−0.0136	0.0200	−0.0046	−0.0038	−0.0076
	RMSE	0.1073	0.1322	0.1261	0.1136	0.1203	0.1567	0.1484
2SGM	Bias	0.0085	−0.0101	−0.0197	−0.0467	0.1252	−0.0040	0.0190
	RMSE	0.0895	0.1167	0.1408	0.1003	0.1574	0.1693	0.2165
HMLE	Bias	0.0027	0.0019	−0.0082	0.0247	−0.0080	−0.0038	−0.0146
	RMSE	0.0867	0.1099	0.1057	0.0922	0.0933	0.1676	0.2048
HIV	Bias	0.0009	−0.0040	−0.0259	0.0682	0.0066	−0.0022	−0.0085
	RMSE	0.0881	0.1133	0.4538	0.1109	0.0952	0.1581	0.1951
BHIV	Bias	0.0014	−0.0030	−0.0122	0.0240	0.0054	−0.0024	−0.0083
	RMSE	0.0875	0.1118	0.1146	0.0929	0.0929	0.1522	0.1792
HGMM	Bias	−0.0007	−0.0019	−0.0124	0.0256	0.0112	−0.0089	−0.0163
	RMSE	0.0859	0.1112	0.1067	0.0932	0.0932	0.1552	0.1787

1: iid $N(0,1)$; 2: Mixed normal: iid $0.95N(0,1)+0.05N(0,25)$; 3: NC: $0.95N(0,1)+0.05t_1$; 4: Extreme contamination: $\epsilon_1 \sim N(150,1)$; 5: Extreme contamination: $\epsilon_1, \epsilon_2 \sim N(150,1)$; 6: Laplace$\cdot x_{i2,n}$: Laplace distribution scaled by x_{i2}; and 7: exp(1) $\cdot\, x_{i2,n}^2$: exp(1) scaled by $x_{i2,n}^2$. BIV: Best IV; GMM: GMM in Lin and Lee (2010); QIV: quantile IV in Su and Yang (2011); 2SGM: Two-stage generalized M-estimator in Wagenvoort and Waldmann (2002); HMLE: Huber MLE in Tho et al. (2023); HIV: Huber IV; BHIV: Best Huber IV; and HGMM: Huber GMM.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

TABLE 2. Estimation results ($n = 400$)

	Dist.	1: $N(0,1)$	2: Mixed normal	3: NC	4: Extreme I	5: Extreme II	6: Laplace$\cdot x_{i2}$	7: exp(1)$\cdot x_{i2}^2$
					λ			
QMLE	Bias	−0.0054	−0.0065	−0.0105	0.0454	0.0583	−0.0059	−0.0115
	RMSE	0.0366	0.0452	0.0527	0.0458	0.0586	0.0444	0.0443
2SLS	Bias	−0.0010	0.0023	0.0395	−0.0433	0.0660	−0.0004	−0.0004
	RMSE	0.0482	0.0767	0.2119	0.0518	0.0691	0.0725	0.0983
BIV	Bias	−0.0041	−0.0044	−0.0700	−0.0862	0.2212	−0.0035	−0.0166
	RMSE	0.0482	0.0763	2.6794	0.0919	0.2220	0.0733	0.1235
GMM	Bias	−0.0048	−0.0055	0.0168	−0.0022	−0.0023	−0.0041	−0.0019
	RMSE	0.0367	0.0435	0.1781	0.0061	0.0662	0.0442	0.0427
QIV	Bias	0.0005	0.0060	0.0049	−0.0025	0.0040	0.0006	0.0004
	RMSE	0.0596	0.0659	0.0628	0.0457	0.0461	0.0263	0.0170
2SGM	Bias	−0.0049	0.0030	0.0005	0.0168	0.0301	−0.0004	−0.0008
	RMSE	0.0509	0.0623	0.0635	0.0498	0.0544	0.0508	0.0435
HMLE	Bias	−0.0050	−0.0075	−0.0027	−0.0002	0.0301	−0.0157	−0.0076
	RMSE	0.0374	0.0378	0.0398	0.0202	0.0252	0.0377	0.0279
HIV	Bias	−0.0008	0.0010	0.0024	−0.0014	−0.0002	0.0003	−0.0007
	RMSE	0.0483	0.0583	0.0578	0.0318	0.0349	0.0393	0.0291
BHIV	Bias	−0.0039	−0.0019	−0.0014	−0.0034	0.0012	−0.0012	−0.0009
	RMSE	0.0483	0.0564	0.0553	0.0334	0.0345	0.0393	0.0244
HGMM	Bias	−0.0048	−0.0033	−0.0003	0.0008	0.0050	−0.0026	−0.0019
	RMSE	0.0369	0.0378	0.0385	0.0229	0.0246	0.0277	0.0160

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

TABLE 2. Continued

	Dist.	1: $N(0,1)$	2: Mixed normal	3: NC	4: Extreme I	5: Extreme II	6: Laplace$\cdot x_{i2}$	7: exp(1)$\cdot x_{i2}^2$
					β_2			
QMLE	Bias	−0.0006	−0.0006	−0.0232	0.6453	−0.1054	0.0024	−0.0017
	RMSE	0.0493	0.0757	0.9757	0.6471	0.1162	0.1294	0.2262
2SLS	Bias	0.0003	−0.0033	−0.0280	0.6487	−0.1055	0.0010	−0.0036
	RMSE	0.0494	0.0762	0.9928	0.6506	0.1163	0.1294	0.2267
BIV	Bias	−0.0005	−0.0025	−0.0883	0.6508	−0.1031	0.0093	−0.0033
	RMSE	0.0494	0.0762	2.7902	0.6627	0.1139	0.1596	0.2274
GMM	Bias	−0.0012	−0.0024	−0.0302	0.6466	−0.1069	0.0009	−0.0031
	RMSE	0.0495	0.0757	1.0201	0.6484	0.1177	0.1293	0.2260
QIV	Bias	0.0001	−0.0026	−0.0046	0.0054	−0.0016	−0.0021	0.0013
	RMSE	0.0611	0.0667	0.0657	0.0523	0.0618	0.0785	0.0831
2SGM	Bias	0.0020	−0.0005	−0.0024	0.0044	−0.0033	0.0002	0.0021
	RMSE	0.0506	0.0588	0.0548	0.0510	0.0506	0.0837	0.0948
HMLE	Bias	0.0003	0.0002	−0.0016	0.0074	−0.0009	−0.0005	0.0038
	RMSE	0.0505	0.0574	0.0537	0.0509	0.0504	0.0868	0.0976
HIV	Bias	0.0008	−0.0014	−0.0023	0.0080	−0.0004	−0.0025	0.0052
	RMSE	0.0494	0.0577	0.0555	0.0511	0.0509	0.0786	0.0860
BHIV	Bias	0.0005	−0.0012	−0.0018	0.0073	−0.0011	−0.0023	0.0056
	RMSE	0.0494	0.0577	0.0538	0.0510	0.0506	0.0776	0.0819
HGMM	Bias	−0.0012	−0.0013	−0.0025	0.0065	−0.0018	−0.0036	0.0040
	RMSE	0.0496	0.0576	0.0537	0.0507	0.0509	0.0779	0.0823

1: iid $N(0,1)$; 2: Mixed normal: iid $0.95N(0,1)+0.05N(0,25)$; 3: NC: $0.95N(0,1)+0.05t_1$; 4: Extreme contamination: $\epsilon_1 \sim N(150,1)$; 5: Extreme contamination: $\epsilon_1, \epsilon_2 \sim N(150,1)$; 6: Laplace$\cdot x_{i2,n}$: Laplace distribution scaled by x_{i2}; and 7: exp(1) $\cdot\, x_{i2,n}^2$: exp(1) scaled by $x_{i2,n}^2$. BIV: Best IV; GMM: GMM in Lin and Lee (2010); QIV: quantile IV in Su and Yang (2011); 2SGM: Two-stage generalized M-estimator in Wagenvoort and Waldmann (2002); HMLE: Huber MLE in Tho et al. (2023); HIV: Huber IV; BHIV: Best Huber IV; and HGMM: Huber GMM.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

TABLE 3. Relative efficiency of Huber estimators with respect to traditional estimators

Distributions			1: $N(0,1)$	2: Mixed normal	3: NC	4: Extreme I	5: Extreme II	6: Laplace$\cdot x_{i2}$	7: exp(1)$\cdot x_{i2}^2$
	λ	HIV/2SLS	0.8641	1.6486	6.503	105.6	23.22	2.839	7.389
		BHIV/BIV	0.9348	2.0843	527.3	3.2×10^5	541.9	5.223	77.35
		HGMM/GMM	0.9717	1.3407	49.5	167.3	140.1	2.866	1
$n = 100$	β_2	HIV/2SLS	0.9455	1.7550	391.6	356.6	16.11	4.968	6.450
		BHIV/BIV	0.9571	1.8193	3,899	11,165	435.0	7.250	8.993
		HGMM/GMM	0.9930	1.7729	5,337	491.2	25.85	2.700	8.116
	β_3	HIV/2SLS	0.9213	1.8000	349.1	233.3	52.22	3.075	12.10
		BHIV/BIV	0.9501	1.8250	5,395	963.9	109.1	3.169	21.58
		HGMM/GMM	0.9766	1.8102	7,403	245.7	71.99	3.050	18.03
	λ	HIV/2SLS	0.9168	1.7596	9.6	32.91	2.981	2.916	7.792
		BHIV/BIV	0.9453	1.9344	46.7	60.49	80.38	4.193	22.25
		HGMM/GMM	0.9860	1.3142	9.3	6.950	9.621	2.778	15.63
$n = 200$	β_2	HIV/2SLS	0.9477	1.8413	628.8	1.466	1.408	1.180	8.298
		BHIV/BIV	0.9541	1.8472	2,421	33.08	141.0	2.707	9.293
		HGMM/GMM	0.9980	1.8342	4,658	81.53	29.91	2.616	9.243
	β_3	HIV/2SLS	0.9464	1.7991	628.9	52.05	4.773	2.940	9.109
		BHIV/BIV	0.9608	1.7961	392.1	38.19	3.306	3.094	13.91
		HGMM/GMM	0.9778	1.7746	396.3	100.2	2.878	3.071	14.39

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

TABLE 3. Continued

Distributions			1: $N(0,1)$	2: Mixed normal	3: NC	4: Extreme I	5: Extreme II	6: Laplace$\cdot x_{i2}$	7: exp(1)$\cdot x_{i2}^2$
	λ	HIV/2SLS	0.9979	1.7346	13.44	2.653	3.920	3.403	11.41
		BHIV/BIV	0.9979	1.8295	2,344	7.571	41.41	3.479	25.62
		HGMM/GMM	0.9892	1.3237	21.40	0.071	7.242	2.546	7.122
$n = 400$	β_2	HIV/2SLS	0.9960	1.7440	319.7	162.1	5.221	2.710	6.949
		BHIV/BIV	1	1.7446	2,689	168.8	5.067	4.230	7.709
		HGMM/GMM	0.9960	1.7279	360.5	163.6	5.347	2.755	7.541
	β_3	HIV/2SLS	0.9961	1.7443	219.6	6.590	36.56	4.230	19.82
		BHIV/BIV	0.9961	1.7345	9,100	5.502	30.21	4.359	27.52
		HGMM/GMM	0.9961	1.7136	274.4	7.185	40.20	4.258	26.86

1: iid $N(0,1)$; 2: Mixed normal: iid $0.95N(0,1)+0.05N(0,25)$; 3: NC: iid $0.95N(0,1)+0.05t_1$; 4: Extreme contamination: $\epsilon_1 \sim N(150,1)$; 5: Extreme contamination: $\epsilon_1, \epsilon_2 \sim N(150,1)$; 6: Laplace$\cdot x_{i2,n}$: Laplace distribution scaled by x_{i2}; and 7: exp(1) $\cdot\, x_{i2,n}^2$: exp(1) scaled by $x_{i2,n}^2$. BIV: Best IV; GMM: GMM in Lin and Lee (2010): HIV: Huber IV; BHIV: Best Huber IV; HGMM: Huber GMM. Relative efficiency=$MSE(\hat{\theta}_{traditional})/MSE(\hat{\theta}_{Huber})$.

UHII by using temperature difference between impervious surfaces and other types of land use to measure urban heat island effect. They find a similar result: a one-unit increase in the UHII is associated with a 3.91% decrease in home values in Beijing, China. However, their estimation procedure does not address the issue of long-tailed feature and conditional heteroskedasticity of housing price data and may result in larger standard errors if that is the case. To accommodate this issue, we adopt the HGMM approach proposed in this article to examine the urban heat island effect on housing prices in Beijing.

The variable of interest is the UHII proposed by Lin et al. (2022). Control variables include whether a community is located in a top ten primary school district, housing age, distance to the nearest park and subway, normalized difference built-up index (NDBI), and normalized difference vegetation index (NDVI). The detailed definitions and summary statistics are detailed in Table 4. Because housing prices are spatially correlated in nature (Irwin, Jeanty, and Partridge, 2014), we employ an SAR model to control for spatial correlation among housing prices. For the spatial weights matrix $W_n \equiv (w_{ij})_{n\times n}$, $w_{ij} = \frac{1}{4}$ if j is among the four nearest neighbors of i, $w_{ij} = 0$ otherwise. The estimation results are summarized in Table 5. To ascertain the robustness of our estimation findings, we explore two alternative specifications for the spatial weights matrix: (1) $w_{ij} = 1/6$ if j is among the six nearest neighbors of i, and $w_{ij} = 0$ otherwise and (2) $w_{ij} \propto d_{ij}^{-5}$ with the spatial weights matrix row normalized. In the second case, the spatial weights are a decreasing function of distances between communities. These additional estimation results are summarized in Tables F.1 and F.2 in the Supplementary Material.

To demonstrate the importance of the HGMM approach, we compare it with two traditional estimation methods for the SAR model: QMLE and the traditional GMM approach in Lin and Lee (2010). We also perform a naive OLS estimation. From Table 5 and Tables F.1 and F.2 in the Supplementary Material, we observe both conditional heteroskedasticity and a long-tailed distribution in the error terms:

1. There exists conditional heteroskedasticity. For every method, we regress $\hat{\epsilon}_{i,n}^2$ on $x_{i,n}$, where $\hat{\epsilon}_{i,n}$ is the residual, and calculate the test statistic nR^2 to test whether there exists conditional heteroskedasticity. And the test statistics are always greater than 370, and the corresponding p-values are 0. So, there is convincing evidence that there exists conditional heteroskedasticity.
2. The error terms follow a long-tailed distribution. In most results, the residuals exhibit a kurtosis of approximately 9; in one case, it is 69.9. So, the kurtosis of the residuals is clearly greater than 3 (the kurtosis of the normal distribution), which indicates that the disturbances are long-tailed or there exist outliers.

Given the above observations, because both the QMLE and the Huber MLE for the SAR model assume i.i.d. errors and the GMM estimator in Lin and Lee (2010) is sensitive to long-tailed distributions as indicated by simulation studies, the HGMM estimation is suitable for this dataset. As in Section 4, we set $P_n = \tilde{G}_n - \text{diag}(\tilde{G}_n)$,

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

TABLE 4. Definitions and statistical summary of UHI data

Variable name	Definition	Data sources	Mean	Std. Dev.
Price	Annual average price per square meter of the apartments in a community.($)	Beijing Municipal Commission of Housing and Urban–Rural Development	10,653.51	4,471.63
UHII	Urban heat island intensity	Calculated from Landsat TM 8 satellite images	8.674	2.231
School	Equals 1 if the community is located within the top ten primary schools, otherwise 0	Beijing Municipal Education Commission	0.014	0.118
Age	Home transaction year minus built year	Beijing Municipal Commission of Housing and Urban–Rural Development	15.606	5.771
Park	Distance to the nearest park in meters	Calculated using ArcGIS; Beijing Municipal Forestry and Parks Bureau	1.274	0.988
Subway	Distance to the nearest subway station in kilometers	Calculated using ArcGIS; Beijing Traffic Management Bureau	2.237	5.509
NDBI	Normalized difference built-up index, measuring building density	Calculated from Landsat TM 8 satellite images	−0.085	0.074
NDVI	Normalized difference vegetation index, measuring greenness of green vegetation	Calculated from Landsat TM 8 satellite images	0.346	0.154
No. of observations (communities)		6,028		

Table 5. Estimation results under W_{4n}

	OLS	QMLE	GMM	HGMM
λ		0.5145***	0.5410***	0.6297***
		(0.0113)	(0.0170)	(0.0126)
UHII	−0.0131***	−0.0116***	−0.0116***	−0.0082***
	(0.0020)	(0.0016)	(0.0017)	(0.0012)
School	0.1404***	0.0820***	0.0789***	0.0548***
	(0.0190)	(0.0238)	(0.0144)	(0.0131)
Park	−0.0421***	−0.0230***	−0.0213***	−0.0147***
	(0.0052)	(0.0032)	(0.0039)	(0.0031)
Age	0.0035***	0.0028***	0.0027***	0.0002
	(0.0006)	(0.0005)	(0.0005)	(0.0004)
Subway	−0.4388***	−0.2214***	−0.2323***	−0.1709***
	(0.0697)	(0.0196)	(0.0375)	(0.0262)
NDBI990	−0.0985**	−0.0349	−0.0290	−0.0226
	(0.0460)	(0.0373)	(0.0382)	(0.0264)
NDVI210	0.1164***	0.0964***	0.0968***	0.0580***
	(0.0222)	(0.0192)	(0.0195)	(0.0134)
Kurtosis	8.7991	9.0931	9.1535	9.5490
nR^2	439.4	409.3	412.8	377.2
n		6,028		
W_{4n}		Constructed by the nearest four neighbors		

∗∗ and ∗∗∗: Significant at levels 5% and 1%. Standard deviation: Heteroskedasticity-consistent estimator.

where $\tilde{G}_n = W_n(I_n - \tilde{\lambda}W_n)^{-1}$ and $\tilde{\lambda}$ is a consistent estimator for λ_0.[18] The corresponding estimators are denoted as GMM and HGMM, respectively. The value of M is determined as in Section 2.3. From Table 5, the estimates of the spatial interaction coefficient λ in columns 2–4 are all significant and range between 0.5 and 0.63, indicating a strong spatial correlation in housing prices. The absolute values of the naive OLS coefficients are bigger than other estimates in general. So, the effects of UHII and the control variables are probably overestimated if we ignore the spatial correlation in housing prices. Moreover, the standard errors of all the estimates from HGMM are smaller than those of GMM, indicating our HGMM approach is more robust and precise than traditional GMM estimators for this dataset. For UHII in the SAR model, all estimation methods yield significant estimates at the 1% level, showing that higher UHII leads to lower housing prices. The QMLE and the traditional GMM estimates are both −0.0116, while the

[18]For the GMM estimator, we use the 2SLS estimate for $\tilde{\lambda}$; for the HGMM estimator, we use the HIV estimate for $\tilde{\lambda}$.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

HGMM estimate is −0.0082. It seems that QMLE and the traditional GMM tend to overestimate the effects of the urban heat island.

When considering the six nearest neighbors setting for W_n, the estimation results (Table F.1 in the Supplementary Material) are similar. However, when considering the setting $w_{ij} \propto d_{ij}^{-5}$ for W_n, from Table F.2 in the Supplementary Material, compared to the HGMM estimation, the traditional GMM estimation drastically underestimates the spatial interaction coefficient. It suggests that the first two setups for W_n are more reliable.

6. CONCLUSION

In this article, we introduce two computationally convenient estimators for the SAR model that are robust to both long-tailed disturbances and conditional heteroskedasticity: the HIV estimator and the HGMM estimator. Unlike the Huber MLE proposed by Tho et al. (2023), our estimators are also robust to conditional heteroskedasticity. They are not only robust to long-tailed disturbances but also incur minimal efficiency loss compared to traditional estimators for the SAR model under short-tailed disturbances. We establish their large sample properties by extending the empirical process results on non-differentiable objective functions in Pollard (1985)to allow heterogeneous and mutually correlated data. We show that our estimators are consistent under regularity conditions, even when the expectation of the error terms does not exist. In addition, we discuss the breakdown point of our estimators and show that the influence functions of our estimators are bounded. We also extend our estimation methods to the SARARE model, whose error terms are spatially correlated. Simulation studies confirm that these estimators outperform the corresponding traditional estimators when error terms are long-tailed or there exist outliers. Finally, we apply the HGMM estimation to study the effect of urban heat island on housing prices in Beijing. Given the conditional heteroskedasticity and the long-tailed distribution of the error terms, the HGMM estimation is well-suited for this dataset. Our study provides more reliable evidence that the urban heat island effects contribute to a reduction in housing prices in big cities.

We have made our MATLAB code publicly available to support replication and further empirical applications. The code is accessible at: https://github.com/Xingbai2004/Huber-GMM.

SUPPLEMENTARY MATERIAL

Liu, T., Xu, X., Lee, L.-F., & Mei, Y. (2026). Supplement to "Robust estimation for the spatial autoregressive model," Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466626100346.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

FUNDING STATEMENT

T.L. gratefully acknowledges the financial support of the Chinese National Natural Science Fund of China (CNNSF) (Grant Nos. 71903163 and 72033008). X.X. gratefully acknowledges the financial support of the CNNSF (Grant Nos. 72073110, 72473118, and 72333001), Basic Scientific Center Project 71988101 of CNNSF, and the 111 Project (B13028). L.-F.L. gratefully acknowledges the financial support of CNNSF (Grant No. 72333002).

COMPETING INTEREST STATEMENT

The authors declare that no competing interests exist.

REFERENCES

Acemoglu, D. *et al.*, (2019). Democracy does cause growth. *Journal of political economy*, 127(1), 47–100.

Andrews, D. W. (1992). Generic uniform convergence. *Econometric Theory*, 8, 241–257.

Baskaran, T. (2014). Identifying local tax mimicking with Administrative borders and a policy reform. *Journal of Public Economics*, 118, 41–51.

Bickel, P. J. (1975). One-step Huber estimates in the linear model. *Journal of the American Statistical Association*, 70(350), 428–434.

Britos, G. M., & Ojeda, S. M. (2019). Robust estimation for spatial autoregressive processes based on bounded innovation propagation representations. *Computational Statistics*, 34, 1315–1335.

Calvó-Armengol, A., Patacchini, E., & Zenou, Y. (2009). Peer effects and social networks in education. *The Review of Economic Studies*, 76(4), 1239–1267.

Cressie, N. (1993). *Statistics for spatial data*. John Wiley & Sons.

Fan, J., Li, R., Zhang, C. H., & Zou, H. (2020). *Statistical foundations of data science*. Chapman and Hall/CRC.

Folland, G. B. (1999). *Real analysis: Modern techniques and their applications*. John Wiley & Sons.

Gallant, A. R., & White, H. (1988). *A unified theory of estimation and inference for nonlinear dynamic models*. Basil Blackwell.

Gelfand, A. E., Diggle, P., Guttorp, P., & Fuentes, M. (2010). *Handbook of spatial statistics*. CRC press.

Genton, M. G. (2003). Breakdown-point for spatially and temporally correlated observations. In R. Dutter, P. Filzmoser, U. Gather, & P. J. Rousseeuw (Eds.), *Developments in robust statistics* (pp. 148–159). Physica-Verlag HD.

Genton, M. G., & Lucas, A. (2003). Comprehensive definitions of breakdown points for independent and dependent observations. *Journal of the Royal Statistical Society: Series B*, 65, 81–94.

Hampel, F. R. (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, 42(6), 1887–1896.

Hsu, A., Sheriff, G., Chakraborty, T., & Manya, D. (2021). Disproportionate exposure to urban heat island intensity across major US cities. *Nature Communications*, 12, 2721.

Huber, P. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1), 73–101.

Huber, P. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Fifth berkeley symposium on mathematical statistics and probability* (pp. 221–233) Lucien M. Le Cam, Jerzy Neyman, University of California Press.

Huber, P. & Ronchetti, E. (2009). Robust Statistics (2nd ed.). John Wiley & Sons, Inc.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

Irwin, G. E., Jeanty, P. W., & Partridge, D. M. (2014). Amenity values versus land constraints: The spatial effects of natural landscape features on housing values. *Land Economics*, 90(1), 61–78.

Jenish, N., & Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, 150, 86–98.

Jenish, N., & Prucha, I. R. (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics*, 170, 178–190.

Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The 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(2), 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(2), 219–257.

Kelejian, H. H., & Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157, 53–67.

Kim, T. H., & Muller, C. (2007). Two-stage Huber estimation. *Journal of Statistical Planning and Inference*, 137(2), 405–418.

Klaiber, H. A., Abbott, J. K., & Smith, V. K. (2017). Some like it (less) hot: Extracting tradeoff measures for physically coupled amenities. *Journal of the Association of Environmental & Resource Economists*, 4(4), 1053–1079.

Kuersteiner, G. M., & Prucha, I. R. (2013). Limit theory for panel data models with cross sectional dependence and sequential exogeneity. *Journal of Econometrics*, 174(2), 107–126.

Kuersteiner, G. M., & Prucha, I. R. (2020). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica*, 88(5), 2109–2146.

Lee, L. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, 22(4), 307–335.

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

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

Lee, L., Yang, C., & Yu, J. (2023). QML and efficient GMM estimation of spatial autoregressive models with dominant (popular) units. *Journal of Business & Economic Statistics*, 41(2), 550–562.

Lin, L., Meng, L., Mei, Y., Zhang, W., Liu, H., & Xiang, W. (2022). Spatial-temporal patterns of summer urban islands and their economic implications in Beijing. *Environmental Science and Pollution Research*. https://doi.org/10.1007/s11356-021-18029-6

Lin, X., & Lee, L. (2010). GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics*, 157(1), 34–52.

Liu, T., & Lee, L. (2019). A likelihood ratio test for spatial model selection. *Journal of Econometrics*, 213, 434–458.

Liu, T., & Lee, L. (2023). Estimation and selection of high order spatial autoregressive models with adaptive elastic net. Working paper.

Lockwood, B., & Migali, G. (2009). Did the single market cause competition in excise taxes? Evidence from EU countries. *The Economic Journal*, 119(536), 406–429.

Lyytikäinen, T. (2012). Tax competition among local governments: Evidence from a property tax reform in Finland. *Journal of Public Economics*, 96, 584–595.

Mallows, C. L. (1975). On some topics in robustness. Bell Telephone Laboratories, Murray Hill, NJ: Technical Memorandum.

Maronna, R. A., Martin, R. D., Yohai, V. J., & Salibián-Barrera, M. (2019). *Robust statistics: Theory and methods (with R)*. John Wiley & Sons.

Mohajerani, L., Bakaric, J., & Jeffrey-Bailey, T. (2017). The urban heat island effect, its causes, and mitigation, with reference to the thermal properties of asphalt concrete. *Journal of Environmental Management*, 197, 522–538.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

Pakes, A., & Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5), 1027–1057.
Phillips, P. C., & Jiang, L. (2025). Cross section curve autoregression: The unit root case. Manuscript.
Plantinga, A. J., Détang-Dessendre, C., Hunt, G. L., & Piguet, V. (2013). Housing prices and inter-urban migration. *Regional Science and Urban Economics*, 43(2), 296–306.
Pollard, D. (1985). New ways to prove central limit theorems. *Econometric Theory*, 1(3), 295–313.
Pötscher, B. M., & Prucha, I. (1997). *Dynamic nonlinear econometric models: Asymptotic theory*. Springer Science & Business Media.
Qu, X., & Lee, L. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, 184, 209–232.
Sølvsten, M. (2020). Robust estimation with many instruments. *Journal of Econometrics*, 214, 495–512.
Su, L., & Yang, Z. (2011). Instrumental variable quantile estimation of spatial autoregressive models. Working paper.
Tho, Z. Y., Ding, D., Hui, F. K., Welsh, A. H., & Zou, T. (2023). On the robust estimation of spatial autoregressive models. *Econometrics and Statistics* in press.
Wagenvoort, R., & Waldmann, R. (2002). On B-robust instrumental variable estimation of the linear model with panel data. *Journal of Econometrics*, 106, 297–324.
Xiao, X., Xu, X., & Zhong, W. (2023). Huber estimation for the network autoregressive model. *Statistics & Probability Letters*, 203, 109917.
Xu, X., & Lee, L. (2015a). A spatial autoregressive model with a nonlinear transformation of the dependent variable. *Journal of Econometrics*, 186, 1–18.
Xu, X., & Lee, L. (2015b). Maximum likelihood estimation of a spatial autoregressive Tobit model. *Journal of Econometrics*, 188, 264–280.
Xu, X., & Lee, L. (2018). Sieve maximum likelihood estimation of the spatial autoregressive Tobit model. *Journal of Econometrics*, 203, 96–112.
Yildirim, V., & Mert Kantar, Y. (2020). Robust estimation approach for spatial error model. *Journal of Statistical Computation and Simulation*, 90(9), 1618–1638.
Zhong, S., Qian, Y., Zhao, C., Leung, R., Wang, H., Yang, B., Fan, J., Yan, H., Yang, X.-Q., & Liu, D. (2017). Urbanization-induced urban heat island and aerosol effects on climate extremes in the Yangtze River Delta region of China. *Atmospheric Chemistry and Physics*, 17, 5439–5457.

Appendixes

A. SOME LEMMAS

The proofs of the following lemmas are collected in Section D of the Supplementary Material.

LEMMA A.1. *Suppose that* $\mathrm{E}|X|^{\gamma} < \infty$ *for some constant* $\gamma > 0$ *and* $M > 0$ *is a constant. Let* $p \geq \max(1, \gamma)$ *be a constant. Then,*

$$\|\min(M, |X|)\|_{L^p} \leq M^{1-\gamma/p} \left(\mathrm{E}|X|^{\gamma}\right)^{1/p}.$$

LEMMA A.2. $\gamma \in [1,2)$ *is a constant. Suppose that* $\{v_{i,n} : i \in D_n\}_{n\in\mathbb{N}}$ *is uniformly* L_γ*-NED on* $\{\epsilon_{i,n} : i \in D_n\}_{n\in\mathbb{N}}$*:* $||v_{i,n} - \mathrm{E}[v_{i,n}|\mathcal{F}_{i,n}(s)]||_{L^\gamma} \leq \psi(s)$ *for all i, n, and some non-increasing function* $\psi(\cdot)$ *satisfying* $\lim_{s\to\infty} \psi(s) = 0$*, where* $\mathcal{F}_{i,n}(s) \equiv \sigma(\{\epsilon_{j,n} : d_{ij} \leq s\})$ *is the sub-*σ*-field generated by those* $\epsilon_{j,n}$*'s with* $d_{ij} \leq s$*.* $F(\cdot)$ *is an* $\mathbb{R}$*-valued function*

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

bounded by a constant M and is a Lipschitz function with a Lipschitz constant L. Then, $\|F(v_{i,n}) - \mathrm{E}[v_{i,n}|\mathcal{F}_{i,n}(s)]\|_{L^2} \leq (2M)^{1-\gamma/2} L^{\gamma/2} \psi(s)^{\gamma/2}$ *for all i and n.*

LEMMA A.3. *Suppose that* $\{v_{i,n} : i \in D_n\}_{n\in\mathbb{N}}$ *satisfies the conditions in Lemma A.2. Let* $A_n = (a_{ij,n})$ *be a sequence of non-stochastic* $n \times n$ *matrices and* $y_{i,n} \equiv \sum_{j=1}^{n} a_{ij,n} v_{j,n}$. *Then, under Assumption 1, for any two positive numbers* $\bar{s} < s$ *and for any i and n,*

$$\|y_{i,n} - \mathrm{E}[y_{i,n}|\mathcal{F}_{i,n}(s)]\|_{L^\gamma} \leq 2\|A_n\|_\infty \psi(s-\bar{s}) + \|v\|_{L^\gamma} \sum_{j\in D_n : d_{ij}\geq \bar{s}} |a_{ij,n}|,$$

where $\|v\|_{L^\gamma} \equiv \sup_{i,n} \|v_{i,n}\|_{L^\gamma}$.

LEMMA A.4. *For all real numbers* a_1, a_2, b_1, b_2, $|\max(a_1,a_2) - \max(b_1,b_2)| \leq \max(|a_1-b_1|, |a_2-b_2|) \leq |a_1-b_1| + |a_2-b_2|$.

LEMMA A.5. *Let Y and Z be two random variables defined on a probability space* $(\Omega, \mathcal{F}, \Pr)$ *and* $\mathcal{G}$ *is a sub-*σ*-field of* $\mathcal{F}$. *If* $|Y| \leq M$, $\|Z\|_{L^r} \leq C$ *for some* $r > 2$, $\|Y - \mathrm{E}[Y|\mathcal{G}]\|_{L^2} \leq \psi_Y$, *and* $\|Z - \mathrm{E}[Z|\mathcal{G}]\|_{L^2} \leq \psi_Z$, *then*

$$\|YZ - \mathrm{E}[YZ|\mathcal{G}]\|_{L^2} \leq M\psi_Z + 2^{(3r-2)/(2r-2)} C M^{r/(2r-2)} \psi_Y^{(r-2)/(2r-2)}.$$

Remark. Lemma A.5 is a generalization of Lemma A.2 in Xu and Lee (2015b). The above Lemma A.5 requires Y to be bounded, while Lemma A.2 in Xu and Lee (2015b) allows both Y and Z to be unbounded. But the conclusion in the above Lemma A.5 is stronger. We employ the above lemma in this article, rather than that in Xu and Lee (2015b), to relax some moment conditions in our assumptions.

LEMMA A.6. *Denote* $S_n \equiv I_n - \lambda_0 W_n$ *and* $G_n \equiv W_n S_n^{-1}$.
(1) Under Assumptions 1, 3, and 4(i), for all $s \geq 0$,

$$\sup_{n, i\in D_n} \sum_{j\in D_n : d_{ij}\geq s} |(S_n^{-1})_{ij}| \leq \zeta^{s/\bar{d}_0}/(1-\zeta),$$

$$\sup_{n, i\in D_n} \sum_{j\in D_n : d_{ij}\geq s} |(G_n)_{ij}| \leq \left(\sup_n \|W_n\|_\infty\right) \zeta^{s/\bar{d}_0}/(1-\zeta).$$

(2) Under Assumptions 1, 3, and 4(ii), whenever $s > 0$,

$$\sup_{n, i\in D_n} \sum_{j\in D_n : d_{ij}\geq s} |(S_n^{-1})_{ij}| \leq C s^{-(\alpha-d)} [\max(\ln s, 1)]^{\alpha-d}, \tag{A.1}$$

$$\sup_{n, i\in D_n} \sum_{j\in D_n : d_{ij}\geq s} |G_{ij,n}| \leq |\lambda_0^{-1}| C s^{-(\alpha-d)} [\max(\ln s, 1)]^{\alpha-d} \text{ whenever } \lambda_0 \neq 0,$$

for some constant C that depends only on ζ *and* λ_0.

LEMMA A.7. *Suppose that e is a continuous random variable with a continuous density function* $f(e)$ *supported on some interval on* $\mathbb{R}$. *Define* $g(a,b) \equiv \mathrm{E}[\psi(ae+b)h(e)]$, *where*

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

$a \neq 0$ and $\psi(\cdot)$ is defined in Eq. (2.2) and $h(\cdot)$ is a continuous function. Denote $\dot{\psi}(x) = 1(|x| \leq M)$, and it is equal to the derivative of ψ almost everywhere. Then,

(1) $\frac{\partial g(a,b)}{\partial a} = \mathrm{E}[\dot{\psi}(ae+b)eh(e)]$;

(2) $\frac{\partial g(a,b)}{\partial b} = \mathrm{E}[\dot{\psi}(ae+b)h(e)]$;

(3) $\frac{\partial \mathrm{E}[\dot{\psi}(ae+b)h(e)]}{\partial a} = \frac{-M-b}{|a|a} h\left(\frac{-M-b}{a}\right) f\left(\frac{-M-b}{a}\right) - \frac{M-b}{|a|a} h\left(\frac{M-b}{a}\right) f\left(\frac{M-b}{a}\right)$;

(4) $\frac{\partial \mathrm{E}[\dot{\psi}(ae+b)h(e)]}{\partial b} = \frac{1}{|a|} h\left(\frac{-M-b}{a}\right) f\left(\frac{-M-b}{a}\right) - \frac{1}{|a|} h\left(\frac{M-b}{a}\right) f\left(\frac{M-b}{a}\right)$;

(5) $\frac{\partial^2 g(a,b)}{\partial a \partial b} = -\frac{M+b}{|a|a} h\left(\frac{-M-b}{a}\right) f\left(\frac{-M-b}{a}\right) - \frac{M-b}{|a|a} h\left(\frac{M-b}{a}\right) f\left(\frac{M-b}{a}\right)$;

(6) $\frac{\partial^2 g(a,b)}{\partial b^2} = \frac{1}{|a|} h\left(\frac{-M-b}{a}\right) f\left(\frac{-M-b}{a}\right) - \frac{1}{|a|} h\left(\frac{M-b}{a}\right) f\left(\frac{M-b}{a}\right)$.

Especially, when $h(e) \equiv e$, $\frac{\partial g(a,b)}{\partial b} = F(\frac{M-b}{a}) - F(\frac{-M-b}{a})$.

LEMMA A.8. *(1) Under Assumptions 2–4, $\sup_{i,n} \left(||q_{i,n}||_{L^\gamma} + ||y_{i,n}||_{L^\gamma} + ||z_{i,n}||_{L^\gamma}\right) < \infty$, where $q_{i,n}$ can be $x_{i,n}$, $w_{i\cdot,n}X_n$, $w_{i\cdot,n}W_nX_n$, or $G_{i\cdot,n}X_n\beta_0$.*

(2) Under Assumptions 1–3 and 4(1), $\{y_{i,n}\}_{i\in D_n}$, $\{w_{i\cdot,n}Y_n\}_{i\in D_n}$, and $\{G_{i\cdot,n}X_n\beta_0\}_{i\in D_n}$ are uniformly and geometrically L_γ-NED on $\{x_{i,n}, \epsilon_{i,n}\}$, i.e., there exists a constant C_y that depends on none of i, $\bar{d}_0$, n, or ζ, such that for all $s > 0$, $\sup_{i,n} ||u_{i,n} - \mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(s)]||_{L^\gamma} \leq C_y \zeta^{s/\bar{d}_0}$, where $u_{i,n}$ represents $y_{i,n}$, $w_{i\cdot,n}Y_n$, or $G_{i\cdot,n}X_n\beta_0$ and $\mathcal{F}_{i,n}(s) \equiv \sigma(\{x_{j,n}, \epsilon_{j,n} : d_{ij} \leq s\})$.

Under Assumptions 1–3 and 4(1), $\{w_{i\cdot,n}X_n\}_{i\in D_n}$ and $\{w_{i\cdot,n}W_nX_n\}_{i\in D_n}$ are uniformly L_γ-NED on $\{x_{i,n}\}$ such that $\sup_{i,n} ||w_{i\cdot,n}X_n - \mathrm{E}[w_{i\cdot,n}X_n|x_{j,n}, d_{ij} \leq s]||_{L^\gamma} = 0$ when $s \geq \bar{d}_0$ and $\sup_{i,n} ||w_{i\cdot,n}W_nX_n - \mathrm{E}[w_{i\cdot,n}W_nX_n|x_{j,n}, d_{ij} \leq s]||_{L^\gamma} = 0$ when $s \geq 2\bar{d}_0$.

(3) Under Assumptions 1–3 and 4(2), $\{y_{i,n}\}$, $\{w_{i\cdot,n}Y_n\}_{i\in D_n}$, and $\{G_{i\cdot,n}X_n\beta_0\}_{i\in D_n}$ are uniformly L_2-NED on $\{x_{i,n}, \epsilon_{i,n}\}$: there exists a constant C_y that depends on neither i nor n, such that for all $s \geq 1$, $\sup_{i,n} ||u_{i,n} - \mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(s)]||_{L^\gamma} \leq C_y s^{-(\alpha-d)} [\max(1, \ln s)]^{\alpha-d}$, where $u_{i,n}$ represents $y_{i,n}$, $w_{i\cdot,n}Y_n$, or $G_{i\cdot,n}X_n\beta_0$.

Under Assumptions 1–3 and 4(2), $\{w_{i\cdot,n}X_n\}_{i\in D_n}$ and $\{w_{i\cdot,n}W_nX_n\}_{i\in D_n}$ are uniformly L_γ-NED on $\{x_{i,n}\}$, i.e., when $s > 0$, $\sup_{i,n} ||w_{i\cdot,n}X_n - \mathrm{E}[w_{i\cdot,n}X_n|x_{j,n}, d_{ij} \leq s)]||_{L^\gamma} \leq C_y s^{-(\alpha-d)}$ and $\sup_{i,n} ||w_{i\cdot,n}W_nX_n - \mathrm{E}[w_{i\cdot,n}W_nX_n|x_{j,n}, d_{ij} \leq s)]||_{L^\gamma} \leq C_y s^{-(\alpha-d)}$.[19]

(4) Under Assumptions 1–3 and 4(1), $\{\psi(\epsilon_{i,n}(\theta)) : i \in D_n\}$ is L_2-NED on $\{x_{i,n}, \epsilon_{i,n} : i \in D_n\}$ with L_2-NED coefficient $\lesssim \zeta^{\bar{\gamma}s/2\bar{d}_0}$ uniformly in $\theta \in \Theta$, where $\bar{\gamma} = \min(2, \gamma)$; under Assumptions 1–3 and 4(2), $\{\psi(\epsilon_{i,n}(\theta)) : i \in D_n\}$ is L^2-NED on $\{x_{i,n}, \epsilon_{i,n} : i \in D_n\}$ with L_2-NED coefficient $\lesssim s^{-\bar{\gamma}(\alpha-d)/2} [\max(\ln s, 1)]^{\bar{\gamma}(\alpha-d)/2}$ uniformly in $\theta \in \Theta$.

LEMMA A.9. *Under Assumptions 1–4, $P_{kn} = A_n - \mathrm{diag}(A_n)$, where $A_n \in \{W_n, W_n^2, S_n^{-1}, G_n\}$ satisfies $\sup_n ||P_{kn}||_\infty < \infty$ and the following two conditions:*

(i) Under Assumption 4(i), $\{P_{i\cdot,kn}\psi(\epsilon_n(\theta)) : i \in D_n\}$ is uniformly bounded and uniformly (in i, n, and θ) L_2-NED on $\{x_{i,n}\epsilon_{i,n} : i \in D_n\}$ with NED coefficient $\lesssim \zeta^{\bar{\gamma}s/(2\bar{d}_0)}$ when $A_n \in \{W_n, W_n^2\}$ and $\lesssim \zeta^{\bar{\gamma}s/(4\bar{d}_0)}$ when $A_n \in \{S_n^{-1}, G_n\}$, where $P_{i\cdot,kn}$ is the i-th row of P_{kn} and $\bar{\gamma} = \min(2, \gamma)$.

[19] In (3), when s is large, the NED coefficient is of order $s^{-(\alpha-d)}(\ln s)^{\alpha-d}$. But since $\ln 1 = 0$, following the suggestion of one anonymous referee, we change $\ln(2s)$ into $\max(\ln s, 1)$, which does not change the order.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

(ii) Under Assumption 4(ii), $\{P_{i\cdot,kn}\psi(\epsilon_n(\theta)) : i \in D_n\}$ *is uniformly bounded and uniformly (in* i*,* n*, and* θ*)* L_2*-NED on* $\{x_{i,n}\epsilon_{i,n} : i \in D_n\}$ *with NED coefficient* $\lesssim s^{-\bar{\gamma}(\alpha-d)/2}[\max(1, \ln s)]^{\bar{\gamma}(\alpha-d)/2}$.

LEMMA A.10. *X is a random variable, $\mathcal{G}$ is a sub-σ-field, f is a Borel measurable function, $p \geq 1$ is a constant, and $||f(X)||_{L^p} < \infty$. Then,*

$$||f(X) - \mathrm{E}(f(X)|\mathcal{G})||_{L^p} \leq 2||f(X) - f(\mathrm{E}(X|\mathcal{G}))||_{L^p}.$$

If $\mathcal{G}_1$ is a sub-σ-field of $\mathcal{G}_2$, then $||f(X) - \mathrm{E}(f(X)|\mathcal{G}_1)||_{L^p} \leq 2||f(X) - f(\mathrm{E}(X|\mathcal{G}_2))||_{L^p}$.

COROLLARY A.1. *If f is Lipschitz function with Lipschitz constant L, and random field $\{Y_{i,n} : i \in D_n\}$ is uniformly L_p-NED for some $p \geq 1$: $||Y_{i,n} - \mathrm{E}[Y_{i,n}|\mathcal{F}_{i,n}(s)]||_{L^p} \leq \psi(s)$, then $||f(Y_{i,n}) - \mathrm{E}[f(Y_{i,n})|\mathcal{F}_{i,n}(s)]||_{L^p} \leq L\psi(s)$.*

B. SOME RESULTS FOR STOCHASTIC EQUICONTINUITY AND ASYMPTOTIC DISTRIBUTION

Here, we generalize a lemma for stochastic equicontinuity (SEC) in Pollard (1985) into non-identically distributed triangular arrays, which might have cross-sectional correlation. This idea originated from Huber (1967). On a probability space, $X_n \equiv \{X_{i,n} : i = 1, \ldots, n\}$ is a set of random vectors supported on $\mathcal{X}^n$. $X_{i,n}$'s might be independent, time series, panel data, spatially correlated random vectors, or network data. Suppose the parameter space Θ is a bounded convex subset in the d-dimensional euclidean space $\mathbb{R}^d$. And we consider the infinity norm for parameters, i.e., $||\theta||_\infty = \max_{k=1,\ldots,d}|\theta_d|$. Because a ball under the infinity norm turns out to be a cube in $\mathbb{R}^d$, it is more convenient to construct brackets. But other norms also work, as all norms on $\mathbb{R}^d$ are equivalent. Without loss of generality and for simplicity, we can assume the true parameter $\theta_0 = 0$, as in Huber (1967) and Pollard (1985). For any $r > 0$, denote the cube $B(\theta, r) \equiv \{\bar{\theta} \in \mathbb{R}^d : ||\bar{\theta} - \theta||_\infty \leq r\}$ and $B(r) \equiv B(0, r)$. For each i, n, and θ, we have a random function $f_{i,n}(\theta) \equiv f_{i,n}(X_n, \theta) : \mathcal{X}^n \times \Theta \to \mathbb{R}$. For any θ, $\nu_n f(\theta) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^n [f_{i,n}(\theta) - \mathrm{E}f_{i,n}(\theta)]$ denotes an empirical process. As the proof is similar to that in Pollard (1985) we omit it here.[20]

PROPOSITION B.1. *$\mathcal{F} \equiv \{f_{i,n}(\theta) : \mathcal{X} \times \Theta \to \mathbb{R} | \theta \in \Theta, i = 1, \ldots, n, n \in \mathbb{N}\}$ is a set of functions with parameter $\theta \in \Theta$. For any cube $B(\theta, \delta) \subset \Theta$, $\{f_{i,n}(\bar{\theta}) \in \mathcal{F} | \bar{\theta} \in B(\theta, \delta)\}$ can be covered by a bracket $[\underline{f}_{i,n}(\theta, \delta), \bar{f}_{i,n}(\theta, \delta)]$, i.e., for any $\bar{\theta} \in B(\theta, \delta)$ and any i and n, $\underline{f}_{i,n}(\theta, \delta) \leq f_{i,n}(\bar{\theta}) \leq \bar{f}_{i,n}(\theta, \delta)$. All brackets $[\underline{f}_{i,n}(\theta, \delta), \bar{f}_{i,n}(\theta, \delta)]$ satisfy*

$$\sup_{i,n} \mathrm{E}[\bar{f}_{i,n}(\theta, \delta) - \underline{f}_{i,n}(\theta, \delta)] \leq C_0\delta \qquad \textbf{(B.1)}$$

[20] As mentioned in Pollard (1985), the proof of Lemma 4 in Pollard (1985) depends little on the assumption of independent sampling. The condition Eq. (B.2) can be used to replace the condition ($P\sup_{|t|\leq R}|f(\cdot, t)|^2 \to 0$ as $R \to 0$) and the independence condition in Lemma 4 in Pollard (1985). Thus, we will not repeat the proof here.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

for some constant C_0 that does not depend on i, n, θ, or δ, and

$$R(r) \equiv \sup_n \sup_{\theta\in\Theta,\delta>0:||\theta||_\infty+\delta\le r} \mathrm{E}\left\{\left[v_n\bar{f}(\theta,\delta)\right]^2+\left[v_n\underline{f}(\theta,\delta)\right]^2\right\}\to 0 \text{ as } r\downarrow 0, \tag{B.2}$$

where $v_n\bar{f}(\theta,\delta)\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^n[\bar{f}_{i,n}(\theta,\delta)-\mathrm{E}\bar{f}_{i,n}(\theta,\delta)]$ *and* $v_n\underline{f}(\theta,\delta)\equiv\frac{1}{\sqrt{n}}\sum_{i=1}^n[\underline{f}_{i,n}(\theta,\delta)-\mathrm{E}\underline{f}_{i,n}(\theta,\delta)]$. *If* $R(r_0)<\infty$, *then*

$$\mathrm{plim}_{n\to\infty}\sup_{\theta\in B(r_0)}\frac{|v_nf(\theta)|}{1+\sqrt{n}||\theta||_\infty}=0.$$

The next proposition is a direct generalization of Theorem 3.3 in Pakes and Pollard (1989). Its proof is very similar to that for Theorem 3.3 in Pakes and Pollard (1989), and is therefore omitted.

PROPOSITION B.2. *Let $\tilde{G}_n(\theta)$ be an l-dimensional vector-valued stochastic function of a parameter $\theta\in\mathbb{R}^K$ and $\bar{G}_n(\theta)=(\bar{G}_{1,n}(\theta),\ldots,\bar{G}_{l,n}(\theta))'$ be an l-dimensional vector-valued non-stochastic function. $\tilde{W}_n$ is an $l\times l$ positive-definite matrix and the estimator $\hat{\theta}$ satisfies $\tilde{G}_n(\hat{\theta})'\tilde{W}_n\tilde{G}_n(\hat{\theta})\le\min_{\theta\in\Theta}\tilde{G}_n(\theta)'\tilde{W}_n\tilde{G}_n(\theta)+o_p(n^{-1})$. Suppose $\hat{\theta}_n$ is a consistent estimate of the true parameter θ_0, i.e., $\hat{\theta}_n\xrightarrow{p}\theta_0$.*

If (i) $\bar{G}_n(\theta_0)=0$ for all $n\ge K$ and $\liminf_{n\to\infty}|\bar{G}_n(\theta)|>0$ for all $\theta\ne\theta_0$; (ii) $\Gamma_n(\theta)\equiv\partial\bar{G}_n(\theta)/\partial\theta'$ is equicontinuous at θ_0, there exists a positive constant δ_0 such that

$$C_0\equiv\liminf_{n\to\infty}\inf_{\theta_1,\ldots,\theta_l\in B(\theta_0,\delta_0)}\min\mathrm{sv}\left[\frac{\partial\bar{G}_{1,n}(\theta_1)}{\partial\theta},\ldots,\frac{\partial\bar{G}_{l,n}(\theta_l)}{\partial\theta}\right]>0$$

and $\limsup_{n\to\infty}\sup_{\theta\in B(\theta_0,\delta_0)}||\Gamma_n(\theta)||<\infty$, *and* $\Gamma_n(\theta_0)\to\Gamma$, *(iii) for any nonstochastic positive sequence* $\delta_n\to 0$, $\sup_{||\theta-\theta_0||<\delta_n}||\tilde{G}_n(\theta)-\bar{G}_n(\theta)-\tilde{G}_n(\theta_0)||/(n^{-1/2}+||\theta-\theta_0||)=o_p(1)$, *(iv)* $\sqrt{n}\tilde{G}_n(\theta_0)\xrightarrow{d}N(0,V)$, *(v)* $\theta_0\in\Theta^0$, *where Θ^0 denotes the interior of the parameter space $\Theta\subset\mathbb{R}^K$, and (vi) $\tilde{W}_n\xrightarrow{p}W$ for some positive-definite matrix W, then*

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\xrightarrow{d}N(0,(\Gamma'W\Gamma)^{-1}\Gamma'WVW\Gamma(\Gamma'W\Gamma)^{-1}).$$

C. PROOFS OF IMPORTANT RESULTS IN THE MAIN TEXT

C.1. Proofs of Section 3.2

Proof of Theorem 1. Denote $\epsilon_{i,n}(\theta)\equiv y_{i,n}-\lambda w_{i\cdot,n}Y_n-x_{i,n}'\beta$, $z_{i,n}\equiv(w_{i\cdot,n}Y_n,x_{i,n}')'$,

$$\hat{T}_n(\theta)\equiv\left[\frac{1}{n}\sum_{i=1}^n\psi(\epsilon_{i,n}(\theta))q_{i,n}\right]'\Omega_n\left[\frac{1}{n}\sum_{i=1}^n\psi(\epsilon_{i,n}(\theta))q_{i,n}\right],$$

$$T_n(\theta)\equiv\left\{\frac{1}{n}\sum_{i=1}^n\mathrm{E}\left[\psi(\epsilon_{i,n}(\theta))q_{i,n}\right]\right\}'\Omega\left\{\frac{1}{n}\sum_{i=1}^n\mathrm{E}\left[\psi(\epsilon_{i,n}(\theta))q_{i,n}\right]\right\}.$$

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

According to Theorem 3.3 in Gallant and White (1988), we need to verify the uniform convergence in probability condition $\sup_{\theta\in\Theta}|\hat{T}_n(\theta)-T_n(\theta)|\xrightarrow{p}0$ and the identifiable uniqueness of $T_n(\theta)$. □

First, under Assumption 7, to prove $\sup_{\theta\in\Theta}|\hat{T}_n(\theta)-T_n(\theta)|\xrightarrow{p}0$, it suffices to show that $\sup_{\theta\in\Theta}|R_n(\theta)|\xrightarrow{p}0$, where $R_n(\theta)\equiv\frac{1}{n}\sum_{i=1}^{n}\left[\psi(\epsilon_{i,n}(\theta))q_{i,n}-\mathrm{E}\,\psi(\epsilon_{i,n}(\theta))q_{i,n}\right]$. Furthermore, by Theorem 1 in Andrews (1992), it suffices to show that (1) $\{R_n(\theta):n\in\mathbb{N}\}$ is stochastically equicontinuous (SEC) and (2) $R_n(\theta)=o_p(1)$ for any $\theta\in\Theta$.

(1) The SEC of $\{R_n(\theta)\}$: For any two θ_1 and θ_2 in Θ and any $j=1,\ldots,p_q$, because

$$|\psi(a)-\psi(b)|\le\min(2M,|a-b|), \tag{C.1}$$

we have

$$\begin{aligned}\left|\psi(y_{i,n}-z_{i,n}'\theta_1)-\psi(y_{i,n}-z_{i,n}'\theta_2)\right|\cdot|q_{ij,n}|&\le|z_{i,n}'(\theta_1-\theta_2)|\cdot|q_{ij,n}|\\&\le||z_{i,n}||\cdot|q_{ij,n}|\cdot||\theta_1-\theta_2||.\end{aligned} \tag{C.2}$$

By Assumption 6, Lemma A.8, and Proposition 1 in Jenish and Prucha (2009), $\{R_n(\theta):n\in\mathbb{N}\}$ is SEC.

(2) For each $\theta\in\Theta$, $R_n(\theta)\xrightarrow{p}0$: Because both $\{\psi(\epsilon_{i,n}(\theta))\}$ and $\{q_{i,n}\}$ are uniformly bounded and uniformly L^2-NED on $\{x_{i,n},\epsilon_{i,n}\}$[21] (Assumption 6 and Lemma A.8), by Lemma A.5, $\{\psi(\epsilon_{i,n}(\theta))q_{i,n}\}$ is uniformly bounded and uniformly L^2-NED. Then, by the WLLN of NED random fields, $R_n(\theta)\xrightarrow{p}0$.

Second, we will establish the identifiable uniqueness of $T_n(\theta)$. For any $\theta\neq\theta_0$, by Assumption 8 and the formula $|a'\Omega a-b'\Omega b|\ge\min\mathrm{sv}(\Omega)\,\|a-b\|^2$, we have

$$|T_n(\theta)-T_n(\theta_0)|\ge\min\mathrm{sv}(\Omega)\left\|\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\psi(\epsilon_{i,n}(\theta))q_{i,n}-\psi(\epsilon_{i,n}(\theta_0))q_{i,n}\right]\right\|^2>0. \tag{C.3}$$

In addition, by the formula $|a'\Omega a-b'\Omega b|\le\max\mathrm{sv}(\Omega)\,\|a-b\|^2$ and Eq. (C.2),

$$\begin{aligned}|T_n(\theta_1)-T_n(\theta_2)|&\le\max\mathrm{sv}(\Omega)\left\|\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left[\psi(\epsilon_{i,n}(\theta_1))q_{i,n}-\psi(\epsilon_{i,n}(\theta_2))q_{i,n}\right]\right\|^2\\&\lesssim||\theta_1-\theta_2||^2.\end{aligned}$$

So, $\{T_n(\theta):n\in\mathbb{N}\}$ is equicontinuous. Hence, by Lemma 4.1 in Pötscher and Prucha (1997), θ_0 is identifiably unique. ■

Proof of Theorem 2. Denote $G_{k,n}(\theta)\equiv\frac{1}{n}\sum_{i=1}^{n}\psi(\epsilon_{i,n}(\theta))q_{ik,n}$ and $\bar{G}_{k,n}(\theta)\equiv\mathrm{E}\,G_{k,n}(\theta)$. We will prove this theorem by Proposition B.2. Under Assumptions 7–9, conditions (i), (v), and (vi) in Proposition B.2 hold. In the following, we will verify the remaining conditions. □

Condition (ii) in Proposition B.2: With Assumption 10, it remains to prove $\sup_{\theta\in B(\theta_0,\delta_0)}||\Gamma_n(\theta)||<\infty$, which is a direct result of Assumption 6 and Lemma A.8(1).

[21] For simplicity of presentation, we will ignore "$\{x_{i,n},\epsilon_{i,n}:i\in D_n\}$" in the rest of the proof.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

Condition (iii) in Proposition B.2: For any random function $h(X_{i,n},\theta)$, denote the empirical process by $\nu_n\left[h(X_{i,n},\theta)\right] \equiv n^{-1/2}\sum_{i=1}^{n}\left[h(X_{i,n},\theta) - \mathrm{E}\,h(X_{i,n},\theta)\right]$. We verify that $\{\nu_n\{[\psi(\epsilon_{i,n}(\theta)) - \psi(\epsilon_{i,n})]q_{ik,n}\} : n \in \mathbb{N}\}$ satisfies condition (iii) in Proposition B.2 by Proposition B.1. We consider the infinity norm for θ as in Proposition B.1. For any $\theta \in \Theta$ and a small enough $\delta > 0$ such that $B(\theta,\delta) \subset \Theta$, and for any $\bar{\theta} \in B(\theta,\delta)$,

$$\epsilon_{i,n}(\bar{\theta}) = y_{i,n} - z_{i,n}\bar{\theta} = y_{i,n} - z_{i,n}\theta + z_{i,n}(\theta - \bar{\theta}) \leq y_{i,n} - z_{i,n}\theta + ||z_{i,n}||_1\delta.$$

Similarly, $\epsilon_{i,n}(\bar{\theta}) \geq y_{i,n} - z_{i,n}\theta - ||z_{i,n}||_1\delta$. Since $\psi(\cdot)$ is weakly increasing,

$$\left[\psi\left(y_{i,n} - z_{i,n}\theta - ||z_{i,n}||_1\delta\right),\ \psi\left(y_{i,n} - z_{i,n}\theta + ||z_{i,n}||_1\delta\right)\right] \tag{C.4}$$

is a bracket for $\{\psi(\epsilon_{i,n}(\bar{\theta})) : \bar{\theta} \in B(\theta,\delta)\}$. Hence, we can construct a bracket $[\underline{f}(y_{i,n},z_{i,n},\theta,\delta), \bar{f}(y_{i,n},z_{i,n},\theta,\delta)]$ for $\{G_{k,n}(\bar{\theta}) : \bar{\theta} \in B(\theta,\delta)\}$, where

$$\begin{aligned}
\underline{f}(y_{i,n},z_{i,n},\theta,\delta) \equiv{}& \min\left\{\psi(y_{i,n} - z_{i,n}'\theta - ||z_{i,n}||_1\delta)q_{ik,n}, \psi(y_{i,n} - z_{i,n}'\theta + ||z_{i,n}||_1\delta)q_{ik,n}\right\} \\
&- \psi(\epsilon_{i,n})q_{ik,n}, \\
\bar{f}(y_{i,n},z_{i,n},\theta,\delta) \equiv{}& \max\left\{\psi(y_{i,n} - z_{i,n}'\theta - ||z_{i,n}||_1\delta)q_{ik,n}, \psi(y_{i,n} - z_{i,n}'\theta + ||z_{i,n}||_1\delta)q_{ik,n}\right\} \\
&- \psi(\epsilon_{i,n})q_{ik,n}.
\end{aligned}$$

By Eq. (C.1), $\sup_{i,n}\mathrm{E}\left[\bar{f}(y_{i,n},z_{i,n},\theta,\delta) - \underline{f}(y_{i,n},z_{i,n},\theta,\delta)\right] \leq \sup_{i,n}\mathrm{E}\left[2\delta||z_{i,n}||_1 \cdot |q_{ik,n}|\right] \leq C_1\delta$ for some constant $C_1 < \infty$. Thus, Eq. (B.1) in Proposition B.1 is satisfied.

Second, we will verify Eq. (B.2) in Proposition B.1. We only consider the case where $\gamma \in [1,2)$, as the proof when $\gamma \geq 2$ is similar and easier. Fix a small enough constant r such that $B(\theta_0,r) \subset \Theta$. Consider θ and δ satisfying $||\theta - \theta_0||_\infty + \delta < r$. By the NED result in Lemma A.8 and by Lemmas A.2 and A.5, $\{\psi(y_{i,n} - z_{i,n}'\theta - ||z_{i,n}||_1\delta)q_{ik,n} : i \in D_n, n \in \mathbb{N}\}$ is uniformly bounded and uniformly L_2-NED with NED coefficient $\psi_2^{(2)}(s) \lesssim \zeta^{s\gamma/2\bar{d}_0}$ (under Assumption 4(1)) or $\psi_2^{(2)}(s) \lesssim s^{-(\alpha-d)\gamma/2}[\max(\ln s,1)]^{-(\alpha-d)\gamma/2}$ (under Assumption 4(2)). Similarly, the same conclusion holds for $\{\psi(y_{i,n} - z_{i,n}'\theta + ||z_{i,n}||_1\delta)q_{ik,n} : i \in D_n, n \in \mathbb{N}\}$. Hence, by Lemma A.4, $\{\bar{f}(y_{i,n},z_{i,n},\theta,\delta) : i \in D_n, n \in \mathbb{N}\}$ is uniformly bounded and uniformly L_2-NED with NED coefficient $\lesssim \psi_2^{(2)}(s)$. Let $C_q \equiv \sup_{i,n}\left\|q_{i,n}\right\|$. Therefore, by Eq. (C.1),

$$\begin{aligned}
&\left\|\left[\psi(y_{i,n} - z_{i,n}'\theta - ||z_{i,n}||_1\delta) - \psi(\epsilon_{i,n})\right]q_{ik,n}\right\|_{L^2} \\
&\quad\lesssim \left\|\min[2M, ||z_{i,n}||_1(||\theta - \theta_0||_\infty + \delta)]\right\|_{L^2} \leq C_q\left\|\min[2M, r||z_{i,n}||_1]\right\|_{L^2} \lesssim r^{\gamma/2},
\end{aligned} \tag{C.5}$$

where the second inequality is because $||\theta - \theta_0||_\infty + \delta \leq r$, and the last one follows from Lemmas A.1 and A.8. Hence, under Assumptions 2(ii) and 4 and the condition $\alpha > \left[\frac{2}{\min(2,\gamma)} + 1\right]d$, Lemma A.3 in Jenish and Prucha (2012) is applicable:

$$\sup_n \sup_{\theta\in\Theta,\,\delta>0:||\theta||_\infty+\delta\leq r} \mathrm{E}\left[\nu_n\bar{f}(\cdot,\theta,\delta)\right]^2 \lesssim r^{\gamma/2}.$$

Similarly, $\sup_n \sup_{\theta\in\Theta,\,\delta>0:||\theta||_\infty+\delta\leq r} \mathrm{E}\left[\nu_n\underline{f}(\cdot,\theta,\delta)\right]^2 \lesssim r^{\gamma/2}$, and Eq. (B.2) in Proposition B.1 is satisfied.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

Condition (iv) in Proposition B.2: Notice that $\psi(\epsilon_{i,n}(\theta_0))q_{i,n} = \psi(\epsilon_{i,n})q_{i,n}$. We will show that

$$n^{-1/2}\sum_{i=1}^{n}\psi(\epsilon_{i,n})q_{i,n} \xrightarrow{d} N(0,V_0) \tag{C.6}$$

using the CLT for martingale difference arrays in Kuersteiner and Prucha (2013, Thm. 1). Let $a \in \mathbb{R}^q$ be an arbitrary constant vector with $\|a\| = 1$. Define a sequence of σ-algebras $\mathcal{F}_{i,n} = \sigma\{X_n, \epsilon_{j,n}, j = 1,2,\ldots,i-1\}$. Because $\mathrm{E}[\psi(\epsilon_{i,n})|X_n] = 0$ and $\epsilon_{j,n}$'s are independent, $\{n^{-1/2}\psi(\epsilon_{i,n})a'V_0^{-1/2}q_{i,n}, \mathcal{F}_{i,n}, 1 \le i \le n, n \ge 1\}$ is a martingale difference array. By the uniform boundedness of $\{q_{i,n} : i \in D_n\}$ and Assumption 11, conditions (14)–(16) of Theorem 1 in Kuersteiner and Prucha (2013) hold for this martingale difference array. Hence, $n^{-1/2}\sum_{i=1}^{n}\psi(\epsilon_{i,n})a'V_0^{-1/2}q_{i,n} \xrightarrow{d} N(0,1)$. So, Eq. (C.6) follows from the Cramer–Wald device. ■

C.2. Proofs of Section 3.3

Proof of Theorem 3. Denote $\hat{T}_n(\theta) \equiv g_n(\theta)'\Omega_n g_n(\theta)$ and $T_n(\theta) \equiv [\mathrm{E}\, g_n(\theta)]'\Omega_n[\mathrm{E}\, g_n(\theta)]$. By Theorem 3.3 in Gallant and White (1988), to establish the consistency of $\hat{\theta}_{HGMM}$, we need to verify the uniform convergence in probability condition $\sup_{\theta\in\Theta}|\hat{T}_n(\theta) - T_n(\theta)| \xrightarrow{p} 0$ and the identifiable uniqueness of $\{T_n(\theta) : n \in \mathbb{N}\}$. □

(1) Uniform convergence in probability $\sup_{\theta\in\Theta}|\hat{T}_n(\theta) - T_n(\theta)| \xrightarrow{p} 0$:

For the linear Huber moments, the uniform convergence in probability has been established in the proof of Theorem 1. Now, we investigate the uniform convergence of the quadratic Huber moments. Without loss of generality, assume $m = 1$ and denote $P_n \equiv (P_{ij,n}) \equiv P_{1n}$ to simplify notations. Denote $R_n(\theta) \equiv \frac{1}{n}\psi(\epsilon_n(\theta))'P_n\psi(\epsilon_n(\theta))$. By Theorem 1 in Andrews (1992), it suffices to show that (i) $\{R_n(\theta) - \mathrm{E}R_n(\theta) : n \in \mathbb{N}\}$ is SEC and (ii) $R_n(\theta) - \mathrm{E}R_n(\theta) = o_p(1)$ for any $\theta \in \Theta$.

(i) For any $\theta_1, \theta_2 \in \Theta$,

$$\begin{aligned}
&\mathrm{E}\,|R_n(\theta_1) - R_n(\theta_2)| \\
&= \mathrm{E}\left|\frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(P_{ij,n}+P_{ji,n}\right)\left[\psi(\epsilon_{i,n}(\theta_1)) + \psi(\epsilon_{i,n}(\theta_2))\right]\left[\psi(\epsilon_{j,n}(\theta_1)) - \psi(\epsilon_{j,n}(\theta_2))\right]\right| \\
&\le \frac{2M}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|P_{ij,n}+P_{ji,n}\right|\cdot\mathrm{E}\left[||z_{j,n}||\right]\cdot||\theta_1-\theta_2|| = O_p(1)\cdot||\theta_1-\theta_2|| \\
&\le 2M\left(\sup_n||P_n||_\infty\right)\left(\sup_{n,i}\mathrm{E}||z_{j,n}||\right)\cdot||\theta_1-\theta_2|| \lesssim ||\theta_1-\theta_2||,
\end{aligned}$$

where the inequality follows from Eq. (C.1) and $|\psi(\cdot)| \le M$, and the last one is by $\sup_n||P_n||_\infty < \infty$ (Lemma A.9) and $\sup_{n,i}\mathrm{E}||z_{j,n}|| < \infty$ (Proposition A.8). Hence, by Proposition 1 in Jenish and Prucha (2009), $\{R_n(\theta) : n \in \mathbb{N}\}$ is SEC.

Similarly, $\{\mathrm{E}R_n(\theta) : n \in \mathbb{N}\}$ is equicontinuous. So $\{R_n(\theta) - \mathrm{E}R_n(\theta) : n \in \mathbb{N}\}$ is SEC.

(ii) Recall $\psi(\epsilon_n(\theta))'P_{kn}\psi(\epsilon_n(\theta)) = \sum_{i=1}^{n}\psi(\epsilon_{i,n}(\theta))\left[P_{i\cdot,n}\psi(\epsilon_n(\theta))\right]$. By Lemmas A.8 and A.9, $\{P_{i\cdot,n}\psi(\epsilon_n(\theta)) : i \in D_n\}$ is L^2-NED on $\{x_{i,n}\epsilon_{i,n} : i \in D_n\}$. Since $\psi(\cdot)$ is bounded

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

by M and

$$\sup_{n,i}\sup_{\theta\in\Theta}|P_{i\cdot,n}\psi(\epsilon_n(\theta))| \le M\sup_n||P_n||_\infty < \infty,$$

by Lemma A.5, their product $\{\psi(\epsilon_{i,n}(\theta))\sum_{j=1}^n P_{ij,n}\psi(\epsilon_{j,n}(\theta)) : i \in D_n\}$ is uniformly L^2-NED on $\{x_{i,n}\epsilon_{i,n} : i \in D_n\}$ and uniformly bounded. So, by the WLLN of NED random fields, $R_n(\theta) - \mathrm{E}R_n(\theta) = o_p(1)$.

(2) The identifiable uniqueness of $\{T_n(\theta) : n \in \mathbb{N}\}$: Denote $g_n(\theta) = (g_{qn}(\theta)', g_{ln}(\theta)')'$, where $g_{qn}(\theta)$ and $g_{ln}(\theta)$ are the quadratic and linear moments in $g_n(\theta)$, respectively. For any $\theta \neq \theta_0$, by Assumption 8 and the formula $|a'\Omega a - b'\Omega b| \ge \min \mathrm{sv}(\Omega)\,\|a-b\|^2$, as $n \to \infty$,

$$|T_n(\theta) - T_n(\theta_0)| \ge \min \mathrm{sv}(\Omega)\left\|(g_{qn}(\theta)', g_{ln}(\theta)')'\right\|^2 \ge \min \mathrm{sv}(\Omega)\,\|g_{ln}(\theta)\|^2 > 0.$$

Because both $\{\mathrm{E}g_{ln}(\theta) : n \in \mathbb{N}\}$ and $\{\mathrm{E}g_{qn}(\theta) : n \in \mathbb{N}\}$ are equicontinuous (from the proof of Theorem 1 and the proof of part (1) of this theorem), $\{T_n(\theta) : n \in \mathbb{N}\}$ is equicontinuous. Hence, by Lemma 4.1 in Pötscher and Prucha (1997), θ_0 is identifiably unique. ■

Proof of Theorem 4. We prove this theorem using Proposition B.2. Under Assumptions 7, 9, 12, and 13, conditions (i), (v), and (vi) in Proposition B.2 hold. It suffices to show that $g_n(\theta)$ in Eq. (2.6) satisfies the remaining conditions in Proposition B.2. Since we have verified that the linear Huber conditions satisfy these conditions in the proof of Theorem 2, it remains to verify that the quadratic Huber conditions also satisfy these conditions in Proposition B.2. To simplify notation, without loss of generality, suppose there is only one quadratic Huber condition, i.e., $m = 1$, and denote $P_n \equiv P_{1n} \equiv (P_{ij,n})$ and $\tilde{G}_n(\theta) = \frac{1}{n}[\psi(\epsilon_n(\theta))'P_n\psi(\epsilon_n(\theta)) - \psi(\epsilon_n)'P_n\psi(\epsilon_n)]$. The expression for $\Gamma_n(\theta)$ is in Section C of the Supplementary Material. □

Condition (ii) in Proposition B.2: With Assumption 12, it remains to prove $\sup_\theta ||\Gamma_n(\theta)|| < \infty$, which holds because $z_{i,n}$ is uniformly L^1-bounded and $\sup_n ||P_n||_\infty < \infty$.

Condition (iii) in Proposition B.2: We need to show that for any nonstochastic positive sequence $\delta_n \to 0$,

$$\sup_{\theta:\|\theta-\theta_0\|<\delta_n}\frac{\|\tilde{G}_n(\theta) - \mathrm{E}[\tilde{G}_n(\theta)]\|}{n^{-1/2} + \|\theta-\theta_0\|} = o_p(1).$$

We apply Proposition B.1 to verify this SEC condition by constructing a bracket for $\psi(\epsilon_{i,n}(\theta))P_{i\cdot,n}\psi(\epsilon_n(\theta))$. As in Proposition B.1, we consider the infinity norm for the parameter θ in this step. For any $\theta \in \Theta$ and a small enough $\delta > 0$ such that $B(\theta,\delta) \subset \Theta$, from Eq. (C.4),

$$\left[\psi\left(y_{i,n} - z_{i,n}'\theta - ||z_{i,n}||_1\delta\right),\ \psi\left(y_{i,n} - z_{i,n}'\theta + ||z_{i,n}||_1\delta\right)\right] \equiv [L_{i,n}^{(1)}, U_{i,n}^{(1)}] \tag{C.7}$$

is a bracket for $\{\psi(\epsilon_{i,n}(\bar{\theta})) : \bar{\theta} \in B(\theta,\delta)\}$. To shorten formulas, we assume that all entries of P_n are nonnegative.[22] Then, a bracket for $\{P_{i\cdot,n}\psi(\epsilon_n(\bar{\theta})) : \bar{\theta} \in B(\theta,\delta)\}$ is

$$\left[\sum_{j=1}^n P_{ij,n}\psi\left(y_{j,n} - z_{j,n}'\theta - ||z_{j,n}||_1\delta\right),\ \sum_{j=1}^n P_{ij,n}\psi\left(y_{j,n} - z_{j,n}'\theta + ||z_{j,n}||_1\delta\right)\right] \equiv [L_{i,n}^{(2)}, U_{i,n}^{(2)}].$$

[22] When some entries of P_n are negative, the discussion is similar, but formulas are much longer.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

Thus, a bracket for $\left\{\psi(\epsilon_{i,n}(\theta))\left[P_{i\cdot,n}\psi(\epsilon_n(\theta))\right]-\psi(\epsilon_{i,n})\left[P_{i\cdot,n}\psi(\epsilon_n)\right]:\bar{\theta}\in B(\theta_0,\delta)\right\}$ is $[\underline{f}_{i,n}(\theta,\delta),\bar{f}^{i,n}(\theta,\delta)]$, where

$$\underline{f}_{i,n}(\theta,\delta)=\min\left\{L_{i,n}^{(1)}L_{i,n}^{(2)},L_{i,n}^{(1)}U_{i,n}^{(2)},U_{i,n}^{(1)}L_{i,n}^{(2)},U_{i,n}^{(1)}U_{i,n}^{(2)}\right\}-\psi(\epsilon_{i,n})\left[P_{i\cdot,n}\psi(\epsilon_n)\right],$$

$$\text{and }\bar{f}^{i,n}(\theta,\delta)=\max\left\{L_{i,n}^{(1)}L_{i,n}^{(2)},L_{i,n}^{(1)}U_{i,n}^{(2)},U_{i,n}^{(1)}L_{i,n}^{(2)},U_{i,n}^{(1)}U_{i,n}^{(2)}\right\}-\psi(\epsilon_{i,n})\left[P_{i\cdot,n}\psi(\epsilon_n)\right].$$

It remains to verify Eqs. (B.1) and (B.2) in Proposition B.1.

Equation (B.1): It suffices to verify the L^1 norm of the differences of the four terms in $\bar{f}^{i,n}(\theta,\delta)$ and the four terms in $\underline{f}_{i,n}(\theta,\delta)$ are of order δ. Because $C_1\equiv\sup_n||P_n||_\infty<\infty$ and $|\psi(\cdot)|\le M$,

$$\begin{aligned}\sup_{i,n}\mathrm{E}|L^{(1)}U^{(2)}-L^{(1)}L^{(2)}|&\le M\sup_{i,n}\mathrm{E}|U^{(2)}-L^{(2)}|\\&\le M\sup_{i,n}\mathrm{E}\sum_{j=1}^{n}P_{ij,n}\left|\psi\left(y_{j,n}-z_{j,n}'\theta+||z_{j,n}||_1\delta\right)-\psi\left(y_{j,n}-z_{j,n}'\theta-||z_{j,n}||_1\delta\right)\right|\\&\le 2C_1C_2M\cdot\delta,\end{aligned}$$

where the last inequality is by Eq. (C.1) and $C_2\equiv\sup_{j,n}\mathrm{E}||z_{j,n}||_1<\infty$ (Lemma A.8(1)). The other terms can be proved similarly.

Equation (B.2): Let $r>0$ and let $\theta\in\Theta$ and $\delta>0$ satisfy $||\theta-\theta_0||_\infty+\delta<r$. Consider the case when $r\to 0$. By Lemmas A.8 and A.9, all terms (e.g., $L_{i,n}^{(1)}L_{i,n}^{(2)}$) in the expressions of $\bar{f}^{i,n}(\theta,\delta)$ and $\underline{f}_{i,n}(\theta,\delta)$ are uniformly L_2-NED with NED coefficient uniformly (in i, n, θ, and δ) $\lesssim\zeta^{-\bar{\gamma}s/(4\bar{d}_0)}$ (under Assumption 4(1)) or $\lesssim s^{-\bar{\gamma}(\alpha-d)/2}[\max(1,\ln s)]^{\bar{\gamma}(\alpha-d)/2}$ (under Assumption 4(2)). Since both min and max are Lipschitz functions (Lemma A.4), $\{\bar{f}^{i,n}(\theta,\delta):i\in D_n\}$ and $\{\underline{f}_{i,n}(\theta,\delta):i\in D_n\}$ are both uniformly L_2-NED with the same NED coefficient.

Next, we will evaluate $\left\|\bar{f}^{i,n}(\theta,\delta)\right\|_{L^2}$ and $\left\|\underline{f}_{i,n}(\theta,\delta)\right\|_{L^2}$. By a similar argument as for Eq. (C.5),

$$\sup_{i,n}\sup_{\theta\in\Theta,\delta>0:\,||\theta-\theta_0||_\infty+\delta<r}\left\|L_{i,n}^{(1)}-\psi(\epsilon_{i,n})\right\|_{L^2}\lesssim r^{\gamma/2}.$$

The same conclusion holds for $\left\|U_{i,n}^{(1)}-\psi(\epsilon_{i,n})\right\|_{L^2}$, $\left\|L_{i,n}^{(2)}-P_{i\cdot,n}\psi(\epsilon_n)\right\|_{L^2}$, and $\left\|U_{i,n}^{(2)}-P_{i\cdot,n}\psi(\epsilon_n)\right\|_{L^2}$. With these results, direct calculation shows that

$$\sup_{i,n}\sup_{\theta\in\Theta,\delta>0:\,||\theta-\theta_0||_\infty+\delta<r}\left\|L_{i,n}^{(1)}L_{i,n}^{(2)}-\psi(\epsilon_{i,n})\left[P_{i\cdot,n}\psi(\epsilon_n)\right]\right\|_{L^2}\lesssim r^{\gamma/2},$$

where the second inequality is by $|L_{i,n}^{(2)}|\le C_1M$ and $|\psi(\epsilon_n)|\le M$. The same conclusion also holds for the other terms in the expressions of $\bar{f}^{i,n}(\theta,\delta)$ and $\underline{f}_{i,n}(\theta,\delta)$, i.e., $L_{i,n}^{(1)}U_{i,n}^{(2)}$, $U_{i,n}^{(1)}L_{i,n}^{(2)}$, and $U_{i,n}^{(1)}U_{i,n}^{(2)}$. With these conclusions, it follows from $|\min(a,b)|\le|a|+|b|$ and

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

$|\max(a,b)| \leq |a| + |b|$ that

$$\sup_{i,n} \sup_{\theta\in\Theta,\delta>0:\,||\theta-\theta_0||_\infty+\delta<r} \max\left[\left\|f_{-i,n}(\theta,\delta)\right\|_{L^2}, \left\|\bar{f}^{i,n}(\theta,\delta)\right\|_{L^2}\right] \lesssim r^{\gamma/2}.$$

In the above two paragraphs, we have shown that $\{\bar{f}^{i,n}(\theta,\delta) : i \in D_n\}$ and $\{f_{-i,n}(\theta,\delta) : i \in D_n\}$ are uniformly L^2-bounded and uniformly (in i, n, θ, and δ) L_2-NED. Hence, under the condition $\alpha > \left[\frac{2}{\min(2,\gamma)} + 1\right] d$ in this theorem, it follows from Lemma A.3 in Jenish and Prucha (2012) that

$$\sup_{n} \sup_{\theta\in\Theta,\delta>0:\,||\theta-\theta_0||_\infty+\delta<r} \max\left[\left\|\nu[\bar{f}^{i,n}(\theta,\delta)]\right\|_{L^2}, \left\|\nu[f_{-i,n}(\theta,\delta)]\right\|_{L^2}\right] \lesssim r^{\gamma/2},$$

i.e., Eq. (B.2) in Proposition B.1 follows.

So, by Proposition B.1, $\{\tilde{G}_{1n}(\theta) - \tilde{G}_{1n}(\theta_0) : n \in \mathbb{N}\}$ is SEC, i.e., condition (iii) in Proposition B.2 is satisfied.

Condition (iv) (the CLT) in Proposition B.2: Let $a \in \mathbb{R}^q$ be an arbitrary constant vector with $\|a\| = 1$. As in Kelejian and Prucha (2001) and Kuersteiner and Prucha (2020), the linear quadratic form $n^{-1/2}a'V_0^{-1/2}g_n(\theta)$ can be decomposed into the summation of a martingale difference array. And it is easy to verify that this martingale difference array satisfies the conditions of the CLT (Theorem 1) in Kuersteiner and Prucha (2013). Thus, $n^{-1/2}a'V_0^{-1/2}g_n(\theta) \xrightarrow{d} N(0,1)$. So, by the Cramer–Wald device, $\sqrt{n}g_n(\theta_0) \xrightarrow{d} N(0,V_0)$.

Since all conditions in Proposition B.2 are satisfied, the conclusion of this theorem follows. ■

Proof of Theorem 5. (1) We first prove that $\hat{V} \xrightarrow{p} V_0$. To do so, we need to prove that (i) $\frac{1}{n}[(\hat{\Sigma}_n^d)'(P_{in} \odot P_{jn}^s)\hat{\Sigma}_n^d - (\Sigma_n^d)'(P_{in} \odot P_{jn}^s)\Sigma_n^d] \xrightarrow{p} 0$ for all $i,j \in \{1,\ldots,m\}$ and (ii) $\frac{1}{n}[Q_n'\hat{\Sigma}_n Q_n - \mathrm{E}(Q_n'\Sigma_n Q_n)] \xrightarrow{p} 0$. □

(i) To simplify notation, without loss of generality, assume $m = 1$ and denote $P_n \equiv P_{1n}$ and $P_{\Delta n} \equiv P_n \odot P_n^s \equiv (P_{ij,\Delta n})$. Since $(\Sigma_n^d)'(P_{1n} \odot P_{1n}^s)\Sigma_n^d = \sum_{i,j} P_{ij,\Delta n}\mathrm{E}\,\psi(\epsilon_{i,n})^2 \mathrm{E}\,\psi(\epsilon_{j,n})^2$, to prove (i), it suffices to prove $\frac{1}{n}\sum_{i,j} P_{ij,\Delta n}\left[\psi(\epsilon_{i,n})^2\psi(\epsilon_{j,n})^2 - \mathrm{E}\,\psi(\epsilon_{i,n})^2 \mathrm{E}\,\psi(\epsilon_{j,n})^2\right] \xrightarrow{p} 0$, which been proved in Lin and Lee (2010), and

$$\frac{1}{n}\sum_{i,j} P_{ij,\Delta n}\left[\psi(\hat{\epsilon}_{i,n})^2\psi(\hat{\epsilon}_{j,n})^2 - \psi(\epsilon_{i,n})^2\psi(\epsilon_{j,n})^2\right] \xrightarrow{p} 0. \tag{C.8}$$

For Eq. (C.8), since $\hat{\theta} \xrightarrow{p} \theta_0$, it suffices to show that for any non-stochastic sequence $\theta_n \to \theta_0$, we have

$$A_n \equiv \frac{1}{n}\sum_{i,j} P_{ij,\Delta n}\left[\psi(\epsilon_{i,n}(\theta_n))^2\psi(\epsilon_{j,n}(\theta_n))^2 - \psi(\epsilon_{i,n})^2\psi(\epsilon_{j,n})^2\right] \xrightarrow{p} 0. \tag{C.9}$$

To establish Eq. (C.9), it is sufficient to show that $\mathrm{E}A_n \to 0$ and $A_n - \mathrm{E}A_n \xrightarrow{p} 0$, respectively.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

The Proof of $\mathrm{E}A_n \to 0$: As $P_{ii,\Delta n} = 0$ for all i,

$$\begin{aligned}
\mathrm{E}A_n &= \frac{1}{n}\sum_{i,j} P_{ij,\Delta n}\,\mathrm{E}\Big[\psi(\epsilon_{i,n}(\theta_n))^2\psi(\epsilon_{j,n}(\theta_n))^2 - \psi(\epsilon_{i,n})^2\psi(\epsilon_{j,n})^2\Big] \\
&= \frac{1}{n}\sum_{i\neq j} P_{ij,\Delta n}\,\mathrm{E}\Big\{\Big[\psi(\epsilon_{i,n}(\theta_n))^2 - \psi(\epsilon_{i,n})^2 + \psi(\epsilon_{i,n})^2\Big]\cdot \\
&\quad \Big[\psi(\epsilon_{j,n}(\theta_n))^2 - \psi(\epsilon_{j,n})^2 + \psi(\epsilon_{j,n})^2\Big] - \Big[\psi(\epsilon_{i,n})^2\psi(\epsilon_{j,n})^2\Big]\Big\} \\
&= \frac{1}{n}\sum_{i\neq j} P_{ij,\Delta n}\,\mathrm{E}\Big\{\Big[\psi(\epsilon_{i,n}(\theta_n))^2 - \psi(\epsilon_{i,n})^2\Big]\Big[\psi(\epsilon_{j,n}(\theta_n))^2 - \psi(\epsilon_{j,n})^2\Big]\Big\} \\
&\quad + \frac{1}{n}\sum_{i\neq j} P_{ij,\Delta n}\,\mathrm{E}\Big\{\Big[\psi(\epsilon_{i,n}(\theta_n))^2 - \psi(\epsilon_{i,n})^2\Big]\psi(\epsilon_{j,n})^2\Big\} \\
&\quad + \frac{1}{n}\sum_{i\neq j} P_{ij,\Delta n}\,\mathrm{E}\Big\{\psi(\epsilon_{i,n})^2\Big[\psi(\epsilon_{j,n}(\theta_n))^2 - \psi(\epsilon_{j,n})^2\Big]\Big\} = T_1 + T_2 + T_3.
\end{aligned}$$

It remains to show $T_i \to 0$ for $i = 1,2,3$. Their proofs are similar and thus we will only present the proof of $T_1 \to 0$ in the following: As $|\psi(\cdot)| \le M$,

$$\begin{aligned}
|T_1| &\le \frac{M^2}{n}\sum_{i\neq j}\big|P_{ij,\Delta n}\big|\,\mathrm{E}\big|\big[\psi(\epsilon_{i,n}(\theta_n)) - \psi(\epsilon_{i,n})\big]\big[\psi(\epsilon_{i,n}(\theta_n)) + \psi(\epsilon_{i,n})\big]\big| \\
&\le \frac{2M^3}{n}\sum_{i\neq j}\big|P_{ij,\Delta n}\big|\,\mathrm{E}\big|\psi(\epsilon_{i,n}(\theta_n)) - \psi(\epsilon_{i,n})\big| \\
&\le \frac{2M^3}{n}\sum_{i\neq j}\big|P_{ij,\Delta n}\big|\,\mathrm{E}\Big|\min[2M, |z'_{i,n}(\theta_n - \theta_0)|]\Big| \qquad \textbf{(C.10)} \\
&\le \frac{2M^3}{n}\sum_{i\neq j}\big|P_{ij,n}(P_{ij,n} + P_{ji,n})\big|\,\mathrm{E}[2M|z'_{i,n}(\theta_n - \theta_0)|]^{1/2} \\
&\le 2M^3||\theta_n - \theta_0||^{1/2}\cdot\Big(2\sup_n||P_n||_\infty^2\Big)\cdot\sup_{n,i}\frac{1}{2}\mathrm{E}[2M + ||z_{i,n}||] \to 0
\end{aligned}$$

as $n \to \infty$, where the third inequality follows from Eq. (C.1), the last inequality follows from Cauchy's inequality, and the last step is by $\theta_n \to \theta_0$ and Lemma A.8.

The Proof of $A_n - \mathrm{E}A_n \xrightarrow{p} 0$: As

$$A_n \equiv \frac{1}{n}\sum_{i,j} P_{ij,\Delta n}\psi(\epsilon_{i,n}(\theta_n))^2\psi(\epsilon_{j,n}(\theta_n))^2 - \frac{1}{n}\sum_{i,j} P_{ij,\Delta n}\psi(\epsilon_{i,n})^2\psi(\epsilon_{j,n})^2 \equiv A_{1n} - A_{2n},$$

it suffices to show $A_{1n} - \mathrm{E}A_{1n} \xrightarrow{p} 0$ and $A_{2n} - \mathrm{E}A_{2n} \xrightarrow{p} 0$, respectively. As A_{2n} is a special case of A_{1n} (with $\theta_n = \theta_0$), it suffices to show $A_{1n} - \mathrm{E}A_{1n} \xrightarrow{p} 0$. Since $\{\psi(\epsilon_{i,n}(\theta_n)) : i \in D_n\}$ is bounded by M and is uniformly L_2-NED on $\{x_{i,n}, \epsilon_{i,n} : i \in D_n\}$, $\{\psi(\epsilon_{i,n}(\theta_n))^2 : i \in D_n\}$ is also uniformly L_2-NED. By similar argument as in the proof of Lemma A.9,

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

$\{\sum_j P_{ij,\Delta n}\psi(\epsilon_{i,n}(\theta_n))^2\psi(\epsilon_{j,n}(\theta_n))^2 : i \in D_n\}$ is also L_2-NED. Thus, the WLLN for spatial NED is applicable: $A_{1n} - \mathrm{E}A_{1n} \xrightarrow{p} 0$.

(ii) To prove (ii), it suffices to show that for any non-stochastic sequence $\theta_n \to \theta_0$, we have

$$\frac{1}{n}\sum_i \left\{\psi(\epsilon_{i,n}(\theta_n))^2 q_{i,n}q'_{i,n} - \mathrm{E}[\psi(\epsilon_{i,n})^2 q_{i,n}q'_{i,n}]\right\} \xrightarrow{p} 0.$$

In order to prove the above equation, it suffices to show that

$$\frac{1}{n}\sum_i \left\{\psi(\epsilon_{i,n}(\theta_n))^2 q_{i,n}q'_{i,n} - \mathrm{E}[\psi(\epsilon_{i,n}(\theta_n))^2 q_{i,n}q'_{i,n}]\right\} \xrightarrow{p} 0, \tag{C.11}$$

$$\frac{1}{n}\sum_i \left\{\mathrm{E}[\psi(\epsilon_{i,n}(\theta_n))^2 q_{i,n}q'_{i,n}] - \mathrm{E}[\psi(\epsilon_{i,n})^2 q_{i,n}q'_{i,n}]\right\} \to 0. \tag{C.12}$$

The Proof of Eq. (C.11): As $\{q_{i,n} : i \in D_n\}$ is uniformly bounded and L_2-NED dependent (Lemma A.8), by Lemma A.5, every entry of $\{q_{i,n}q'_{i,n} : i \in D_n\}$ is uniformly bounded and L_2-NED dependent. Since $\{\psi(\epsilon_{i,n}(\theta_n))^2 : i \in D_n\}$ is L_2-NED, by Lemma A.5, every entry of $\{\psi(\epsilon_{i,n}(\theta_n))^2 q_{i,n}q'_{i,n} : i \in D_n\}$ is uniformly bounded and L_2-NED. Hence, Eq. (C.11) follows from the WLLN for spatial NED.

The Proof of Eq. (C.12): For any $j,k \in \{1,\ldots,p_q\}$, since $|\psi(\cdot)| \le M$ and $C_0 = \sup_{i,n}\|q_{i,n}\| < \infty$,

$$\begin{aligned}
&\left|\frac{1}{n}\sum_i \left\{\mathrm{E}[\psi(\epsilon_{i,n}(\theta_n))^2 q_{ij,n}q_{ik,n}] - \mathrm{E}[\psi(\epsilon_{i,n})^2 q_{ij,n}q_{ik,n}]\right\}\right| \\
&= \left|\frac{1}{n}\sum_i \mathrm{E}\left\{\left|q_{ij,n}q_{ik,n}\left[\psi(\epsilon_{i,n}(\theta_n)) - \psi(\epsilon_{i,n})\right]\left[\psi(\epsilon_{i,n}(\theta_n)) + \psi(\epsilon_{i,n})\right]\right|\right\}\right| \\
&\le \frac{2MC_0^2}{n}\sum_i \mathrm{E}\left|\psi(\epsilon_{i,n}(\theta_n)) - \psi(\epsilon_{i,n})\right| \\
&\le 2MC_0^2||\theta_n - \theta_0||^{1/2}\cdot\frac{1}{n}\sum_i \mathrm{E}[2M + ||z_{i,n}||] \to 0,
\end{aligned}$$

where the second inequality is obtained similarly as Eq. (C.10).

(2) We will prove $\hat{\Gamma}_n \xrightarrow{p} \Gamma_0$ in this step. Denote $P_{*n} \equiv P^s_{1n} \odot G_n \equiv (P_{ij,*n})$. To show $\hat{\Gamma} \xrightarrow{p} \Gamma_0$, we need to prove the following three equations of convergence: for any non-stochastic sequence $\theta_n \to \theta_0$,

$$\frac{1}{n}\sum_{i,j} P_{ij,*n}\left\{\dot{\psi}(\epsilon_{i,n}(\theta_n))\psi(\epsilon_{j,n}(\theta_n))\epsilon_{j,n}(\theta_n) - \mathrm{E}[\dot{\psi}(\epsilon_{i,n})\psi(\epsilon_{j,n})\epsilon_{j,n}]\right\} \xrightarrow{p} 0, \tag{C.13}$$

$$\frac{1}{n}\sum_i \{\dot{\psi}(\epsilon_{i,n}(\theta_n))(w_{i\cdot,n}Y_n)q_{i,n} - \mathrm{E}[\dot{\psi}(\epsilon_{i,n})(w_{i\cdot,n}Y_n)q_{i,n}]\} \xrightarrow{p} 0, \tag{C.14}$$

$$\frac{1}{n}\sum_i \{\dot{\psi}(\epsilon_{i,n}(\theta_n))q_{i,n}x'_{i,n} - \mathrm{E}[\dot{\psi}(\epsilon_{i,n})q_{i,n}x'_{i,n}]\} \xrightarrow{p} 0. \tag{C.15}$$

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

The Proof of Eq. (C.13): To prove Eq. (C.13), it suffices to prove that

$$\frac{1}{n}\sum_{i,j} P_{ij,*n}\left\{\dot{\psi}(\epsilon_{i,n}(\theta_n))\psi(\epsilon_{j,n}(\theta_n))\epsilon_{j,n}(\theta_n) - \mathrm{E}[\dot{\psi}(\epsilon_{i,n}(\theta_n))\psi(\epsilon_{j,n}(\theta_n))\epsilon_{j,n}(\theta_n)]\right\} \xrightarrow{p} 0, \tag{C.16}$$

$$\frac{1}{n}\sum_{i,j} P_{ij,*n}\left\{\mathrm{E}[\dot{\psi}(\epsilon_{i,n}(\theta_n))\psi(\epsilon_{j,n}(\theta_n))\epsilon_{j,n}(\theta_n)] - \mathrm{E}[\dot{\psi}(\epsilon_{i,n}(\theta_n)\psi(\epsilon_{j,n})\epsilon_{j,n}]\right\} \to 0. \tag{C.17}$$

The Proof of Eq. (C.16): Since $\{\epsilon_{i,n}(\theta_n) : i \in D_n\}$ is uniformly L_γ-NED and $g(\epsilon) \equiv \psi(\epsilon)\epsilon$ is a Lipschitz function, by Corollary A.1, $\{\psi(\epsilon_{i,n}(\theta_n))\epsilon_{i,n}(\theta_n) : i \in D_n\}$ is uniformly L_γ-NED. Then, by Lemma A.3, $\{\sum_j P_{ij,*n}\psi(\epsilon_{j,n}(\theta_n))\epsilon_{j,n}(\theta_n) : i \in D_n\}$ is also uniformly L_γ-NED.

Note that

$$\begin{aligned}
\epsilon_{i,n}(\theta_n) &= y_{i,n} - z_{i,n}\theta = \epsilon_{i,n} + x_{i,n}'(\beta_0 - \beta_n) + (\lambda_0 - \lambda_n)w_{i\cdot,n}Y_n \\
&= \epsilon_{i,n} + (\lambda_0 - \lambda_n)G_{i\cdot,n}(X_n\beta_0 + \epsilon_n) + x_{i,n}'(\beta_0 - \beta_n) \\
&= \epsilon_{i,n}[1 + (\lambda_0 - \lambda_n)G_{ii,n}] + (\lambda_0 - \lambda_n)\left(G_{i\cdot,n}X_n\beta_0 + G_{i,-i,n}\epsilon_{-i,n}\right) + x_{i,n}'(\beta_0 - \beta_n) \\
&= A_{i,n}(\lambda_n)\epsilon_{i,n} + B_{i,n}(\theta_n),
\end{aligned} \tag{C.18}$$

where $G_{i\cdot,n}$ is the i-th row of G_n, $G_{i,-i,n}$ is a $1 \times (n-1)$ vector obtained by deleting the i-th column element in the i-th row of G_n, $\epsilon_{-i,n}$ is obtained by deleting the i-th element of $\epsilon_{i,n}$, $A_{i,n}(\lambda_n) \equiv 1 + (\lambda_0 - \lambda)G_{ii,n}$, $B_{i,n}(\theta) \equiv (\lambda_0 - \lambda)D_{i,n} + x_{i,n}'(\beta_0 - \beta)$, and $D_{i,n} \equiv G_{i\cdot,n}X_n\beta_0 + G_{i,-i,n}\epsilon_{-i,n}$. Since $\sup_n ||G_n||_\infty \equiv C_4 < \infty$ for some constant C_4 and $\theta_n \to \theta_0$, when n is large enough, $\inf_i A_{i,n}(\lambda_n) > 0.5$. And we will only consider this case (n is large enough) in the following. As $\epsilon_{i,n}$'s are independent and $\inf_i A_{i,n}(\lambda_n) > 0.5$, the p.d.f. of $A_{i,n}(\lambda_n)\epsilon_{i,n} + (\lambda_0 - \lambda_n)G_{i,-i,n}\epsilon_{-i,n}$ is equal to

$$\begin{aligned}
&\left(\frac{1}{A_{i,n}(\lambda_n)}f_{in}\left(\frac{\cdot}{A_{i,n}(\lambda_n)}|x_{i,n}\right) * f_{-i,n}(\cdot)\right)(x) \\
&\quad \le \sup\left[\frac{1}{A_{i,n}(\lambda_n)}f_{in}\left(\frac{\cdot}{A_{i,n}(\lambda_n)}|x_{i,n}\right)\right] \le 2\sup_{i,n,\epsilon,x}|f_{in}(\epsilon|x)| < \infty,
\end{aligned}$$

where $*$ denotes convolution and $f_{-i,n}(\cdot)$ denotes the p.d.f. of $(\lambda_0 - \lambda_n)G_{i,-i,n}\epsilon_{-i,n}$, the first inequality is by Young's inequality for convolution (Theorem 8.7 in Folland, 1999, p. 240), and the last one is from Assumption 2. That is to say, the p.d.f.s of $\epsilon_{i,n}(\theta_n)$ are uniformly bounded in i, n, and θ. Because $\{\epsilon_{i,n}(\theta_n) : i \in D_n\}$ is L_γ-NED on $\{x_{i,n}, \epsilon_{i,n} : i \in D_n\}$, by similar argument as in that of Proposition 2 in Xu and Lee (2015b) (replace the first inequality there by Lemma A.10 in this article), $\{\dot{\psi}(\epsilon_{i,n}(\theta_n)) : i \in D_n\}$ is also L_γ-NED on $\{x_{i,n}, \epsilon_{i,n} : i \in D_n\}$. Therefore, it follows from similar arguments as for Lemma A.10 that $A \equiv \{\dot{\psi}(\epsilon_{i,n}(\theta_n))\sum_j P_{ij,n}\psi(\epsilon_{j,n}(\theta_n))\epsilon_{j,n}(\theta_n) : i \in D_n\}$ is also L_γ-NED on $\{x_{i,n}, \epsilon_{i,n} : i \in D_n\}$. Finally, as A is uniformly L^γ-bounded, Eq. (C.16) follows from the WLLN for spatial NED.

The Proof of Eq. (C.17): Since $\sup_n ||P_n||_\infty < \infty$, it suffices to show that

$$\sup_{i,j,n}\left\|\mathrm{E}\left\{\frac{\partial}{\partial\theta}\mathrm{E}[\dot{\psi}(\epsilon_{i,n}(\theta_n))\psi(\epsilon_{j,n}(\theta_n))\epsilon_{j,n}(\theta_n)|\epsilon_{-i,n}, X_n]\right\}\right\| < \infty. \tag{C.19}$$

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

Using Lemma A.7 repeatedly, we have

$$
\begin{aligned}
&\mathrm{E}\left\{\frac{\partial}{\partial\lambda}\mathrm{E}[\dot\psi(\epsilon_{i,n}(\theta_n))\psi(\epsilon_{j,n}(\theta_n))\epsilon_{j,n}(\theta_n)|\epsilon_{-i,n},X_n]\right\}\\
&\quad=-\mathrm{E}[\dot\psi(\epsilon_{i,n}(\theta_n))\dot\psi(\epsilon_{j,n}(\theta_n))\epsilon_{j,n}(\theta_n)G_{j\cdot,n}(X_n\beta_0+\epsilon_n)]\\
&\qquad-\mathrm{E}[\dot\psi(\epsilon_{i,n}(\theta_n))\psi(\epsilon_{j,n}(\theta_n))G_{j\cdot,n}(X_n\beta_0+\epsilon_n)]\\
&\qquad+\mathrm{E}\left\{\mathrm{E}\left[\frac{-G_{ii,n}}{A_{i,n}(\lambda_n)}\frac{-M-B_{i,n}(\theta_n)}{A_{i,n}(\lambda_n)}\psi(*)\left[\epsilon_{j,n}^{(-i)}(\theta_n)+(\lambda_0-\lambda_n)G_{ji,n}\frac{-M-B_{i,n}(\theta_n)}{A_{i,n}(\lambda)}\right]f_{in}\right.\right.\\
&\qquad\left.\left(\frac{-M-B_{i,n}(\theta_n)}{A_{i,n}(\lambda_n)}\right)\right.\\
&\qquad+\frac{G_{ii,n}}{A_{i,n}(\lambda_n)}\frac{M-B_{i,n}(\theta)}{A_{i,n}(\lambda_n)}\psi(*)\left[\epsilon_{j,n}^{(-i)}(\theta_n)+(\lambda_0-\lambda)G_{ji,n}\frac{M-B_{i,n}(\theta)}{A_{i,n}(\lambda_n)}\right]f_{in}\left(\frac{M-B_{i,n}(\theta)}{A_{i,n}(\lambda_n)}\right)\\
&\qquad-\frac{D_{i,n}}{A_{i,n}(\lambda_n)}\psi(*)\left[\epsilon_{j,n}^{(-i)}(\theta_n)+(\lambda_0-\lambda_n)G_{ji,n}\frac{-M-B_{i,n}(\theta_n)}{A_{i,n}(\lambda_n)}\right]f_{in}\left(\frac{-M-B_{i,n}(\theta_n)}{A_{i,n}(\lambda_n)}\right)\\
&\qquad\left.+\frac{D_{i,n}}{A_{i,n}(\lambda_n)}\psi(*)\left[\epsilon_{j,n}^{(-i)}(\theta_n)+(\lambda_0-\lambda_n)G_{ji,n}\frac{M-B_{i,n}(\theta_n)}{A_{i,n}(\lambda_n)}\right]f_{in}\left(\frac{M-B_{i,n}(\theta_n)}{A_{i,n}(\lambda_n)}\right)\middle|X_n,\epsilon_{-i,n}\right]\right\}\\
&\quad=T_1+T_2+\cdots+T_6,
\end{aligned}
$$

where $*$ represents a term that does not matter for our analysis, $\epsilon_{j,n}^{(-i)}(\theta_n)\equiv\epsilon_{j,n}(\theta_n)-(\lambda_0-\lambda_n)G_{ji,n}\epsilon_{i,n}$ denotes the term obtained by deleting the term $(\lambda_0-\lambda_n)G_{ji,n}\epsilon_{i,n}$ in $\epsilon_{j,n}(\theta_n)$. We will bound $T_1,\ldots,T_6$ one by one in the following:

T_1 and T_2: Since $|\dot\psi(\cdot)|\le 1$, $\sup_n||G_n||<\infty$, $\sup_{j,n,\theta}\left|\dot\psi(\epsilon_{j,n}(\theta))\epsilon_{j,n}(\theta)\right|\le M$, $\sup_{i,n}\mathrm{E}|x'_{i,n}\beta_0+\epsilon_{i,n}|<\infty$, by Minkowski's inequality, we have $\sup_{i,n}\mathrm{E}|T_1|<\infty$ and $\sup_{i,n}\mathrm{E}|T_2|<\infty$.

T_3 and T_4: Since $\sup_{i,n}\sup_{x,\epsilon}|\epsilon f_{in}(\epsilon|x)|<\infty$, $\sup_{i,n}\sup_{x,\epsilon}|\epsilon^2 f_{in}(\epsilon|x)|<\infty$, $|\psi(\cdot)|\le M$, $\sup_{i,j,n}\mathrm{E}|\epsilon_{j,n}^{(-i)}(\theta_n)|<\infty$, $\inf_{i,n}A_{i,n}(\lambda_n)>0.5$, $\sup_{i,n}|G_{ii,n}|\le\sup_n||G_n||_\infty<\infty$, we have $\sup_{i,n}\mathrm{E}|T_3|<\infty$ and $\sup_{i,n}\mathrm{E}|T_4|<\infty$.

T_5 and T_6: As $x_{i,n}$'s and $\epsilon_{i,n}$'s are uniformly L^2-bounded, by the Cauchy–Schwarz inequality, $\sup_{i,n}\mathrm{E}|D_{i,n}\epsilon_{j,n}^{(-i)}(\theta_n)|\le\sup_{i,n}\left\|D_{i,n}\right\|_{L^2}\left\|\epsilon_{j,n}^{(-i)}(\theta_n)\right\|_{L^2}<\infty$. In addition, since $|\psi(\cdot)|\le M$, $\sup_{in}\sup_{x,\epsilon}|\epsilon f_{in}(\epsilon|x)|<\infty$, $\inf_{i,n}A_{i,n}(\lambda_n)>0.5$, and $\sup_n||G_n||_\infty<\infty$, we have $\sup_{i,n}\mathrm{E}|T_5|<\infty$ and $\sup_{i,n}\mathrm{E}|T_6|<\infty$.

In sum, $\sup_{i,j,n}\left|\mathrm{E}\left\{\frac{\partial}{\partial\lambda}\mathrm{E}[\dot\psi(\epsilon_{i,n}(\theta_n))\psi(\epsilon_{j,n}(\theta_n))\epsilon_{j,n}(\theta_n)|\epsilon_{-i,n},X_n]\right\}\right|<\infty$. The proof of

$$
\sup_{i,j,n}\left|\mathrm{E}\left\{\frac{\partial}{\partial\beta_k}\mathrm{E}[\dot\psi(\epsilon_{i,n}(\theta_n))\psi(\epsilon_{j,n}(\theta_n))\epsilon_{j,n}(\theta_n)|\epsilon_{-i,n},X_n]\right\}\right|<\infty
$$

is similar. So, Eq. (C.19) follows, and immediately, Eq. (C.17) holds.

The Proof of Eq. (C.14): To prove Eq. (C.14), it suffices to prove that

$$
\frac{1}{n}\sum_i\{\dot\psi(\epsilon_{i,n}(\theta_n))(w_{i\cdot,n}Y_n)q_{i,n}-\mathrm{E}[\dot\psi(\epsilon_{i,n}(\theta_n))(w_{i\cdot,n}Y_n)q_{i,n}]\}\xrightarrow{p}0, \tag{C.20}
$$

$$
\frac{1}{n}\sum_i\{\mathrm{E}[\dot\psi(\epsilon_{i,n}(\theta_n))(w_{i\cdot,n}Y_n)q_{i,n}]-\mathrm{E}[\dot\psi(\epsilon_{i,n})(w_{i\cdot,n}Y_n)q_{i,n}]\}\xrightarrow{p}0. \tag{C.21}
$$

The Proof of Eq. (C.20): Since $\{w_{i\cdot,n}Y_n:i\in D_n\}$ is uniformly L^γ-bounded and L_γ-NED, $\{q_{i,n}:i\in D_n\}$ is uniformly bounded and L_2-NED, and $\gamma\ge 2$, by Lemma A.5, $\{w_{i\cdot,n}Y_nq_{i,n}:i\in D_n\}$ is uniformly L^γ-bounded and L_2-NED. In addition, because $\{\dot\psi(\epsilon_{i,n}(\theta_n)):i\in D_n\}$

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press

is uniformly bounded and L_γ-NED (already shown in the proof of Eq. (C.16)), by Lemma A.5, $\{\dot{\psi}(\epsilon_{i,n}(\theta_n))w_{i\cdot,n}Y_nq_{i,n} : i \in D_n\}$ is uniformly L^γ-bounded and L_2-NED. Hence, Eq. (C.20) follows from the WLLN for spatial NED.

The proof of Eq. (C.21) is similar to that of Eq. (C.17) and thus omitted. So, Eq. (C.14) holds.

The Proof of Eq. (C.15): It is similar to that of Eq. (C.14) and thus omitted.

Now, we have proved Eqs. C.13–C.15. Thus, $\hat{\Gamma} \xrightarrow{p} \Gamma_0$.

https://doi.org/10.1017/S0266466626100346 Published online by Cambridge University Press