Maximum Likelihood Estimation of a Spatial Autoregressive Tobit Model*

Xingbai Xu and Lung-fei Lee

December 2, 2013

Department of Economics, The Ohio State University, Columbus, OH.

Abstract

This paper examines a Tobit model with spatial autoregressive interactions. We consider the maximum likelihood estimation for this model and analyze asymptotic properties of the estimator based on the spatial near-epoch dependence of the dependent variable process generated from the model structure. We show that the maximum likelihood estimator is consistent and asymptotically normally distributed. Monte Carlo experiments are performed to verify finite sample properties of the estimator.

JEL: C13, C21, C24, C63

Keywords: spatial autoregressive model; Tobit model; censored data; near-epoch dependence; maximum likelihood; asymptotic distribution

*An earlier version of this paper was presented in the Econometric seminar at OSU. We thank the audience for their comments. We also would like to acknowledge the valuable discussions and suggestions we received from Robert DeJong, Nazgul Jenish and Xi Qu.

1 Introduction

The spatial autoregressive (SAR) model, $Y_n = \lambda W_n Y_n + X_n \beta + \epsilon_n$, has been extensively studied in spatial econometrics. Most of the early studies are summarized in Anselin (1988), Anselin and Bera (1998) and LeSage and Pace (2009). The two-stage least squares estimation (2SLS) is explored in Kelejian and Prucha (1998, 1999), while the generalized method of moments (GMM) is studied in Lee (2007). The large sample properties of quasi-maximum likelihood estimation (QML) is considered in Lee (2004).

In recent years, there has been growing interest in the nonlinear SAR models, since for some types of data such as censored or binary data, linear models cannot fully capture their characteristics. Jenish (2012) studies the nonparametric estimation of spatial near-epoch dependent (NED) random fields. An SAR model with a nonlinear transformation of the dependent variable is considered in Xu and Lee (2013). The smoothed maximum score estimation of binary choice panel models with spatial autoregressive errors can be found in Lei (2013). Qu and Lee (2013b) investigate the estimation of an SAR model with an endogenous spatial weights matrix. To study the nonlinear SAR models, some laws of large numbers (LLN) and central limit theorems (CLT) are needed to extend asymptotic properties of extreme estimators of nonlinear models with serial correlation (e.g. Gallant and White 1988) to spatial correlation. Jenish and Prucha (2009, 2012) have made some fundamental contributions to this area. Uniform LLN and CLT for spatial mixing random fields are studied in Jenish and Prucha (2009). Jenish and Prucha (2012) establish the LLN and CLT for spatial NED random fields.

In the microeconometric literature, the Tobit model has been widely studied since Amemiya (1973). Some asymptotic properties of the maximum likelihood estimator (MLE) of the Tobit model are summarized in Amemiya (1985). Recently, an increasing number of studies have introduced spatial correlation into Tobit models. LeSage (2000) and LeSage and Pace (2009) consider the Bayesian estimation of the spatial autoregressive Tobit (SART) model. Marsh and Mittelhammer (2004) study the performance of the generalized maximum entropy estimation of the SAR and the SART models with Monte Carlo simulations. Some studies pertaining to testing the existence of spatial correlation in the SART model can be found in Amaral and Anselin (2011), Qu and Lee (2012) and Qu and Lee (2013a). There are also some empirical studies using Bayesian spatial

Tobit models, such as Autant-Bernard and LeSage (2011), and Donfouet, Jeanty and Malin (2012). Additionally, Flores-Lagunes and Schnier (2012) consider the estimation of sample selection (type II Tobit) models that exhibit spatial dependence in the error terms using GMM.

However, to the best of our knowledge, so far there have been no formal studies on the asymptotic properties of the estimators of the SART model. In this paper, we develop the NED properties of important variables generated by this model. Next, we establish the consistency and asymptotic normality of the MLE via the LLN and CLT developed in Jenish and Prucha (2012).

The structure of this paper is as follows. In Section 2, we introduce the SART model and discuss the model coherency. In Section 3, we derive NED properties of the dependent variable and some other relevant variables generated by the model. In Section 4, the identification of the SART model and the consistency of its MLE are discussed. The asymptotic normality of the estimator is established in Section 5. In Section 6, we study finite sample properties and robustness of the estimator by Monte Carlo experiments. All of the proofs for propositions and theorems are presented in the Appendices.

2 The Spatial Autoregressive Tobit Model

Assume $\{(y_{i,n}, x_{i,n})\}_{i=1}^n$, where $y_{i,n}$ is censored such that $y_{i,n} \geqslant 0$, is the sample we observed. We denote the position of individual (spatial unit) i as $s_i \in R^d$, a point in the d-dimensional Euclidean space. For simplicity of notations, we also use i to represent s_i. As there are interactions among different individuals, we use the weights matrix $W_n = (w_{ij,n})$, an $n \times n$ matrix whose elements are all nonnegative, to represent their direct interactions. If there is a direct interaction between individuals i and j, then $w_{ij,n} \neq 0$, or $w_{ji,n} \neq 0$ or both; zero, otherwise. As usual, a proper normalization has $w_{ii,n} = 0$ for all i.

Let $F(x) = \max(0, x)$. Notice that $F(x)$ is a non-decreasing convex function, and it is also a Lipschitz function: $|F(x_1) - F(x_2)| \leqslant |x_1 - x_2|$. The SART model studied in this paper refers to the following model:

$$y_{i,n} = F(\lambda_0 w_{i\cdot,n} Y_n + x_{i,n}\beta_0 + \epsilon_{i,n}), \tag{2.1}$$

where $w_{i\cdot,n}$ is the i-th row of the weights matrix W_n. If we generalize the notations max and $F(\cdot)$ a little: $\max(0, (x_1, \cdots, x_n)') = F((x_1, \cdots, x_n)') \equiv (\max(0, x_1), \cdots, \max(0, x_n))'$, the model can also be written as

$$Y_n = \max(0, Y_n^*) = \max(0, \lambda_0 W_n Y_n + X_n\beta_0 + \epsilon_n) = F(\lambda_0 W_n Y_n + X_n\beta_0 + \epsilon_n), \tag{2.2}$$

where $Y_n^* \equiv \lambda_0 W_n Y_n + X_n\beta_0 + \epsilon_n$.[1] The model can be derived as a complete information game with each spatial unit (an agent) maximizing its linear-quadratic utility function subject to non-negativity constraints, given the actions of its links (see Ballester, Calvó-Armengol and Zenou, 2006, and Calvó-Armengol, Patacchini and Zenou, 2009). Qu and Lee (2012) assume the reaction function of an agent is linear, which is implied by a linear-quadratic utility function.

Because our model (2.2) is a system of nonlinear equations with censored dependent variables, it is necessary to discuss the conditions for model coherency (Amemiya 1974). Before doing so, we list our assumptions.

Assumption 1. *Individual units in the economy are located or living in a region $D_n \subset D \subset R^d$, where the cardinality of D_n satisfies $\lim_{n\to\infty} |D_n| = \infty$. The distance $d(i,j)$ between any two different individuals i and j is larger than or equal to a specific positive constant, without loss of generality, say, 1.*

We note that the space D can be a space of economic characteristics, a geographical space or a mixture of both economic and physical spaces. Correspondingly, the distance may refer to economic and/or physical distance. The distance in the above assumption can be induced from any norm on R^d. Assumption 1 implies that there cannot be an infinite number of individuals in a bounded space. This assumption has the implication for the increasing domains asymptotic and rules out the scenario of infilled asymptotic.[2] This setting is introduced in Jenish and Prucha (2009 and 2012) on the development of statistical theory for spatial mixing and NED processes.

Assumption 2. $\zeta \equiv \lambda_m \sup_n ||W_n||_\infty < 1$, *where* $\lambda_m = \sup_{\lambda\in\Lambda} |\lambda|$ *with* Λ *being the compact parameter space of* λ *on the real line.*

Assumption 2 is a useful assumption in the spatial econometrics literature. A similar assumption can be found in the linear SAR model (e.g. Kelejian and Prucha, 1998). Assumption 2 is related

[1] The above model is not the only way to model censored dependent variable in the framework of an SAR model. As discussed in Qu and Lee (2012), another way to model Y_n is $Y_n^* = \lambda_0 W_n Y_n^* + X_n\beta_0 + \epsilon_n$ and $y_{i,n} = \max(0, y_{i,n}^*)$. In this paper, we do not discuss this latent dependent variable model.

[2] Under infilled asymptotic, even some popular estimators, such as the least squares and the method of moments may not be consistent, as noted in Lahiri (1996).

to model coherency. Amemiya (1974) discusses the model coherency in the simultaneous equation Tobit model using the method of principal minor. From Theorem 3 in Amemiya (1974), Eq. (2.2) has a unique solution if and only if every principal minor of $I_n - \lambda W_n$ is positive, which is implied by Assumption 2.[3] With Assumption 2, we can also establish the existence and uniqueness of the vector of dependent variables as a solution for the model system by a contraction mapping, as discussed in Qu and Lee (2013a).

In addition to Assumption 2, we need structures for the weights matrix W_n in order to establish that the dependent variable is an NED process.

Assumption 3. *The weights $w_{ij,n}$ in W_n are all non-negative. They satisfy at least one of the following two conditions:*

(1) Only individuals whose distances are less than or equal to some specific constant may affect each other directly. Without loss of generality, we set it as $\bar{d}_0$, which is greater than 1. That is to say, the element $w_{ij,n}$ of the weights matrix W_n can be non-zero only if $d(i,j) \leqslant \bar{d}_0$.

(2) (i) For every n, the number of columns, $w_{\cdot j,n}$, of W_n with $\lambda_m \sum_{i=1}^{n} w_{ij,n} > \zeta$ is less than or equal to some fixed natural number that does not depend on n, denoted as N; (ii) there exists an $\alpha > d$ and a constant C_0 such that $0 \leqslant w_{ij,n} \leqslant C_0/d(i,j)^{\alpha}$.[4]

In Assumption 3, we discuss two different settings of the weights matrix. Assumption 3(1) allows two individuals to have direct interaction only when they are located within a specific distance. In spatial econometrics and statistics, we usually set $w_{ij,n} \neq 0$ only if locations i and j are contiguous. This satisfies Assumption 3(1).

Assumption 3(2) allows the existence of direct interaction even though two locations are far away from each other, but we require the strength of their interaction in terms of $w_{ij,n}$ to decline with $d(i,j)$ in the power α, where $\alpha > d$. As the dimension of the Euclidean space increases, more points are allowed within a sphere with a specific radius. Thus, if the decaying rate of $w_{ij,n}$ does not increase, when d is large enough, it is possible that the effects of individual spatial units are

[3] Under Assumption 2, by spectral radius theorem, for any $h \times h$ principal submatrix W^* of W_n, $\max_i |eig_i(\lambda W^*)| \leqslant ||\lambda W^*||_\infty \leqslant ||\lambda W_n||_\infty \leqslant \zeta$, where $eig_i(\lambda W^*)$ is the i^{th} characteristic root of λW^*. Thus, when $eig_i(\lambda W^*)$ is real, $\lambda_i \equiv eig_i(I_h - \lambda W^*) \geqslant 1-\zeta$; otherwise, its congugate $\overline{\lambda_i}$ is also a characteristic value of $I_h - \lambda W^*$, thus $\lambda_i \overline{\lambda_i} \geqslant (1-\zeta)^2$. Thus, the corresponding principal minor $|I_h - \lambda W^*| \geqslant (1-\zeta)^h > 0$.

[4] Here, we use ζ, which is related to $||W_n||_\infty$, but we do not mean that the column sum and the row sum have some relationship. What we want to express is that, except for (at most) a fixed number of columns, the other columns satisfy $\lambda_m b_f \sup_j \sum_i w_{ij,n} \leqslant \zeta' < 1$. Since $\max(\zeta, \zeta') < 1$, for simplicity of notations, we just mix up ζ and ζ', which will not result in any conflict in the proof.

not neglectable even when their distances from i are large. This is because there could be many units located in a sphere of a larger dimensional space. In this case, the NED property might not be guaranteed.[5] Assumption 3(2)(ii) includes the setting of Assumption 3(1) and the case that $w_{ij,n}$ with an upper bound decreasing geometrically as $d(i,j)$ increases, i.e. $w_{ij,n} \leqslant C\eta^{d(i,j)}$ for some constants $C > 0$ and $0 < \eta < 1$. However, Assumption 3(2)(ii) is insufficient for our purpose. An additional condition, Assumption 3(2)(i), is needed for the column sums of W_n, which is not imposed in Assumption 3(1). Assumption 3(2)(i) states that the cardinal number $|\{j : \lambda_m \sum_{i=1}^{n} w_{ij,n} > \zeta\}| \leqslant N$. That is to say, the total effects of links on each spatial unit, with at most N individuals excluded, can be bounded by ζ. This scenario corresponds to the existence of a limited number of (larger) spatial units which can have larger aggregated effects on other spatial units, even as the total number of spatial units increases.[6] By Lemma A.1 in Jenish and Prucha (2009), with Assumption 1, we have $|\{j : m \leqslant d(i,j) < m+1\}| < Cm^{d-1}$ for some constant $C > 0$. With this lemma, our Assumptions 1, 2 and 3(2) imply that $\lambda_m \sup_n ||W_n||_1 < \infty$ (see the Appendices for the proof of Lemma 1). Furthermore, the column sum of W_n^l, multiplied by λ_m^l decays geometrically as l increases, as stated in the following Lemma 1.

Lemma 1. *Under Assumptions 1, 2 and 3(2),* $\lambda_m^l ||W_n^l||_1 \leqslant lN\Gamma\zeta^{l-1}$, *where* $\Gamma = \lambda_m \sup_n ||W_n||_1 < \infty$.

This lemma has explored a feature in W_n^l so the result is established for $||W_n^l||_1$ as a whole[7] and it will be used in the establishment of NED properties of the log Jacobian determinant term in the log-likelihood function and the derivatives of the log Jacobian determinant. The Taylor expansion is applied to the log Jacobian determinant and its derivatives, and thus W_n^l appears. This is why Lemma 1 is required.

[5] In a network, this would refer to a dense network. With a dense network, MLE of a network SAR model might not even exist (see Smith (2009)).

[6] In a network setting, this rules out the existence of many stars in a network. If there were too many stars, the induced correlations among nodes might be too strong to allow the process to be NED. Strong stars may relate to the existence of strong dependence and spatial correlation generates usually weak dependence as noted in Chudik, Pesaran and Tosetti (2011).

[7] The inequality may not be valid for $(\lambda_m ||W_n||)^l$.

3 Moment and NED Properties of Some Variables

In addition to the existence and uniqueness of the dependent variable, in order to study the asymptotic properties of the MLE (or other estimation methods), some moment and NED properties are needed. We first review the definition and some related properties of NED random fields in Jenish and Prucha (2012) for the convenience of reference.

For any random variable v , $||v||_p = [E|v|^p]^{\frac{1}{p}}$ denotes its L_p-norm. Let $Z = \{Z_{i,n}, i \in D_n, n \geqslant 1\}$ and $\epsilon = \{\epsilon_{i,n}, i \in D_n, n \geqslant 1\}$ be two random fields, where D_n satisfies Assumption 1. Suppose that $||Z_{i,n}||_p < \infty$, where $p \geqslant 1$. $\mathcal{F}_{i,n}(s)$ is denoted as a σ-field generated by the random variables $\epsilon_{j,n}$'s with units j's located within the ball $B_i(s)$ with distance s from unit i. Z is said to be L_p-*near-epoch dependent on* ϵ if $||Z_{i,n} - E(Z_{i,n}|\mathcal{F}_{i,n}(s))||_p \leqslant d_{i,n}\psi(s)$ for some array of finite positive constants $d = \{d_{i,n}, i \in D_n, n \geqslant 1\}$ and for some sequence $\psi(s) \geqslant 0$ with $\lim_{s\to\infty}\psi(s) = 0$. The $d_{i,n}$'s are called *NED scaling factors*. The $\psi(s)$, called the *NED coefficients*, can be non-increasing without loss of generality. The NED random field is *uniform* iff $\sup_n \sup_{i\in D_n} d_{i,n} < \infty$, and it is called *geometric* with NED coefficients $\psi(s)$ iff $\psi(s) = O(\rho^s)$ for some $0 < \rho < 1$. The NED property is kept under summation, product (details are summarized in Lemma A.2) and Lipschitz transformations.

To obtain the NED property, we need to make additional assumptions about the model.

Assumption 4. $\sup_{i,n}|x_{i,n}| < \infty$.

Assumption 4 assumes that the exogenous regressors in $x_{i,n}$ are uniformly bounded. This simplified assumption is for convenience in analysis.

Assumption 5. *For each* n, $\epsilon_{i,n}$*'s are i.i.d.* $(0, \sigma_0^2)$ *double arrays.*

Assumption 5 does not impose normality at this stage. Eventually, we shall assume that the disturbances $\epsilon_{i,n}$'s are normally distributed in order for the likelihood function in Eq. (4.1) to be properly specified so that the ML estimate is consistent. We make the weaker Assumption 5 to highlight some general stochastic structures implied solely by the model (2.1) without a strong distributional assumption. The following moment property of $\{y_{i,n}\}_{i=1}^n$ via Eq. (2.1) is an important one.

Lemma 2. *Under Assumptions 2, 4 and 5, if* $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$*, where* p *is a positive integer, then* $\{y_{i,n}\}_{i=1}^n$ *is uniformly* L_p *bounded, i.e.* $\sup_{i,n} \mathrm{E}|y_{i,n}|^p < \infty$.

Proposition 1. *(1) Under Assumptions 1, 2, 3(1), 4 and 5,* $\{y_{i,n}\}_{i=1}^n$ *is a uniformly and geometrically* L_2*-NED random field on* $\{\epsilon_{i,n}\}$ *with NED coefficient* $(\zeta^{1/\bar{d}_0})^s$.

(2) Under Assumptions 1, 2, 3(2), 4 and 5, $\{y_{i,n}\}_{i=1}^n$ *is a uniformly* L_2*-NED random field on* $\{\epsilon_{i,n}\}$*:* $||y_{i,n} - E[y_{i,n}|\mathcal{F}_{i,n}(s)]||_2 \leqslant C/s^{\alpha-d}$ *for some constant* $C > 0$.

In later analyses, we will often deal with $\{w_{i.,n}Y_n\}_{i=1}^n$. As a corollary, we have the NED property of $\{w_{i.,n}Y_n\}_{i=1}^n$.

Corollary 1. *(1) Under Assumptions 1, 2, 3(1), 4 and 5,* $\{w_{i.,n}Y_n\}_{i=1}^n$ *is a uniformly and geometrically* L_2*-NED random field on* $\{\epsilon_{i,n}\}$ *with NED coefficient* $(\zeta^{1/\bar{d}_0})^s$.

(2) Under Assumptions 1, 2, 3(2), 4 and 5, $\{w_{i.,n}Y_n\}_{i=1}^n$ *is a uniformly* L_2*-NED random field on* $\{\epsilon_{i,n}\}$*:* $||w_{i.,n}Y_n - E[w_{i.,n}Y_n|\mathcal{F}_{i,n}(s)]||_2 \leqslant C/s^{\alpha-d}$ *for some constant* $C > 0$.

Recall $Y_n^* \equiv \lambda_0 W_n Y_n + X_n\beta_0 + \epsilon_n$. For notational simplicity, let $z_{i,n}(\theta) \equiv (y_{i,n} - \lambda w_{i\cdot,n}Y_n - x_{i,n}\beta)/\sigma$. Then the uniform L_p boundedness and uniform NED properties of $\{z_{i,n}(\theta)\}_{i=1}^n$ and $\{y_{i,n}^*\}_{i=1}^n$ are direct consequences of Lemma 2, Proposition 1 and Corollary 1:

Corollary 2. *(1) Under Assumptions 2, 4 and 5, if* $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$*, where* p *is a positive integer, then for each* $\theta \in \Theta$*,* $\{z_{i,n}(\theta)\}_{i=1}^n$ *is uniformly* L_p *bounded.*

(2) Under Assumptions 1, 2, 3(1), 4 and 5, for each $\theta \in \Theta$*,* $\{z_{i,n}(\theta)\}_{i=1}^n$ *is a uniformly and geometrically* L_2*-NED random field on* $\{\epsilon_{i,n}\}$ *with NED coefficient* $(\zeta^{1/\bar{d}_0})^s$.

(3) Under Assumptions 1, 2, 3(2), 4 and 5, for each $\theta \in \Theta$*,* $\{z_{i,n}(\theta)\}_{i=1}^n$ *is a uniformly* L_2*-NED random field on* $\{\epsilon_{i,n}\}$*:* $||z_{i,n}(\theta) - E[z_{i,n}(\theta)|\mathcal{F}_{i,n}(s)]|| \leqslant C/s^{\alpha-d}$ *for some constant* $C > 0$.

Corollary 3. *(1) Under Assumptions 2, 4 and 5, if* $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$*, where* p *is a positive integer, then* $\left\{y_{i,n}^*\right\}_{i=1}^n$ *is also uniformly* L_p *bounded.*

(2) Under Assumptions 1, 2, 3(1), 4 and 5, $\{y_{i,n}^*\}_{i=1}^n$ *is a uniformly and geometrically* L_2*-NED random field on* $\{\epsilon_{i,n}\}$ *with NED coefficient* $(\zeta^{1/\bar{d}_0})^s$.

(3) Under Assumptions 1, 2, 3(2), 4 and 5, $\left\{y_{i,n}^*\right\}_{i=1}^n$ *is a uniformly* L_2*-NED random field on* $\{\epsilon_{i,n}\}$*:* $||y_{i,n}^* - E[y_{i,n}^*|\mathcal{F}_{i,n}(s)]||_2 \leqslant C/s^{\alpha-d}$ *for some constant* $C > 0$.

As we will see later, we not only deal with linear functions of the dependent variables, but also some nonlinear functions of them. To analyze the uniform L_p boundedness of those nonlinear functions, we can bound them and their derivatives by some polynomial functions. With these polynomial bounds and Lemma 2, we can establish the uniform L_p boundedness. With Lemma A.1, which is able to relate relevant polynomial functions to the L_2-NED property of those of basic dependent variables, we obtain Lemma A.5. Lemma A.5, which is a generalization beyond Lipschitz functions to a class of functions, can preserve the NED property of their arguments. Next we establish the NED properties of relevant functions of the model:

Lemma 3. *(1) Under Assumptions 1, 2, 3(1), 4 and 5, if* $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$ *for some integer* $p \geqslant 5$*, then for each* $\theta \in \Theta$*,* $\{z_{i,n}^2(\theta)\}_{i=1}^n$ *is a uniformly and geometrically* L_2*-NED random field on* $\{\epsilon_{i,n}\}$ *with NED coefficient* $[\zeta^{(p-4)/((2p-4)\bar{d}_0)}]^s$.

(2) Under Assumptions 1, 2, 3(2), 4 and 5, if $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$ *for some integer* $p \geqslant 5$*, then for each* $\theta \in \Theta$*,* $\{z_{i,n}^2(\theta)\}_{i=1}^n$ *is a uniformly* L_2*-NED random field on* $\{\epsilon_{i,n}\}$*:* $||z_{i,n}^2(\theta) - \mathrm{E}[z_{i,n}^2(\theta)|\mathcal{F}_{i,n}(s)||_2 \leqslant C(1/s^{\alpha-d})^{(p-4)/(2p-4)}$ *for some constant* $C > 0$.

Since the error terms are normally distributed in the Tobit model, we will deal with the distribution and density functions of the standard normal distribution quite often. Here, it is necessary to discuss the properties of some variables related to the normal distribution. Let $\Phi(\cdot)$ and $\phi(\cdot)$, respectively, be the distribution and density functions of the standard normal random variable.

Lemma 4. *(1) Under Assumptions 2, 4 and 5, if* $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$ *for some integer* $p \geqslant 2$*, then for each* $\theta \in \Theta$*,* $\{\ln \Phi(z_{i,n}(\theta))\}_{i=1}^n$ *is uniformly* $L_{p/2}$ *bounded.*

(2) Under Assumptions 1, 2, 3(1), 4 and 5, if $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$ *for some integer* $p \geqslant 5$*, then for each* $\theta \in \Theta$*,* $\{\ln \Phi(z_{i,n}(\theta))\}_{i=1}^n$ *is also uniformly and geometrically* L_2*-NED on* $\{\epsilon_{i,n}\}$ *with NED coefficient* $[\zeta^{(p-4)/((2p-4)\bar{d}_0)}]^s$.

(3) Under Assumptions 1, 2, 3(2), 4 and 5, if $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$ *for some integer* $p \geqslant 5$*, then for each* $\theta \in \Theta$*,* $\{\ln \Phi(z_{i,n}(\theta))\}_{i=1}^n$ *is also uniformly* L_2*-NED on* $\{\epsilon_{i,n}\}$*: for some constant* $C > 0$*,*

$$||\ln \Phi(z_{i,n}(\theta)) - \mathrm{E}[\ln \Phi(z_{i,n}(\theta))|\mathcal{F}_{i,n}(s)]||_2 \leqslant C(1/s^{\alpha-d})^{(p-4)/(2p-4)}.$$

In the following two lemmas, $\phi(\cdot)/\Phi(\cdot)$ is the inverse Mills ratio, which appears in the log-likelihood function of the model.

Lemma 5. *(1) Under Assumptions 2, 4 and 5, if* $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$ *for some integer* $p \geqslant 1$*, then for each* $\theta \in \Theta$*,* $\{\phi(z_{i,n}(\theta))/\Phi(z_{i,n}(\theta))\}_{i=1}^n$ *is uniformly* L_p *bounded.*

(2) Under Assumptions 1, 2, 3(1), 4 and 5, if $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$ *for some integer* $p \geqslant 7$*, then* $\{\phi(z_{i,n}(\theta))/\Phi(z_{i,n}(\theta))\}_{i=1}^n$ *is uniformly and geometrically* L_2*-NED on* $\{\epsilon_{i,n}\}$ *with NED coefficient* $[\zeta^{(p-6)/((2p-6)\bar{d}_0)}]^s$.

(3) Under Assumptions 1, 2, 3(2), 4 and 5, if $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$ *for some integer* $p \geqslant 7$*, then* $\{\phi(z_{i,n}(\theta))/\Phi(z_{i,n}(\theta))\}_{i=1}^n$ *is uniformly* L_2*-NED on* $\{\epsilon_{i,n}\}$*: for some constant* $C > 0$*,*

$$||\phi(z_{i,n}(\theta))/\Phi(z_{i,n}(\theta)) - \mathrm{E}[\phi(z_{i,n}(\theta))/\Phi(z_{i,n}(\theta))|\mathcal{F}_{i,n}(s)]||_2 \leqslant C(1/s^{\alpha-d})^{(p-6)/(2p-6)}.$$

Lemma 6. *(1) Under Assumptions 2, 4 and 5, if* $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$ *for some integer* $p \geqslant 2$*, then for each* $\theta \in \Theta$*,*$\{\phi(z_{i,n}(\theta))z_{i,n}(\theta)/\Phi(z_{i,n}(\theta))\}_{i=1}^n$ *is uniformly* $L_{p/2}$ *bounded.*

(2) Under Assumptions 1, 2, 3(1), 4 and 5, if $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$ *for some integer* $p \geqslant 9$*, then for each* $\theta \in \Theta$*, then* $\{\phi(z_{i,n}(\theta))z_{i,n}(\theta)/\Phi(z_{i,n}(\theta))\}_{i=1}^n$ *is uniformly and geometrically* L_2*-NED on* $\{\epsilon_{i,n}\}$ *with NED coefficient* $[\zeta^{(p-8)/((2p-8)\bar{d}_0)}]^s$.

(3) Under Assumptions 1, 2, 3(2), 4 and 5, if $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$ *for some integer* $p \geqslant 9$*, then for each* $\theta \in \Theta$*, then* $\{\phi(z_{i,n}(\theta))z_{i,n}(\theta)/\Phi(z_{i,n}(\theta))\}_{i=1}^n$ *is uniformly* L_2*-NED on* $\{\epsilon_{i,n}\}$*: for some constant* $C > 0$*,*

$$||\phi(z_{i,n}(\theta))z_{i,n}(\theta)/\Phi(z_{i,n}(\theta)) - \mathrm{E}[\phi(z_{i,n}(\theta))z_{i,n}(\theta)/\Phi(z_{i,n}(\theta))|\mathcal{F}_{i,n}(s)||_2 \leqslant C(1/s^{\alpha-d})^{(p-8)/(2p-8)}.$$

Another nonlinear transformation of $y_{i,n}$ in the log-likelihood function is $\mathbb{I}(y_{i,n} > 0)$. Notice that it is neither a Lipschitz nor a continuous function of $y_{i,n}$, but its NED property can be established with a boundedness condition on the densities of $\{y^*_{i,n}\}_{i=1}^n$.

Lemma 7. *(1) Under Assumptions 1, 2, 3(1), 4 and 5, if the essential supremums of densities of* $\{y^*_{i,n}\}_{i=1}^n$ *are uniformly bounded in* i *and* n*, then* $\{\mathbb{I}(y_{i,n} > 0)\}_{i=1}^n$ *is a uniformly and geometrically* L_2*-NED random field on* $\{\epsilon_{i,n}\}$ *with NED coefficient* $(\zeta^{1/3\bar{d}_0})^s$.

(2) Under Assumptions 1, 2, 3(2), 4 and 5, if the essential supremums of densities of $\{y^*_{i,n}\}_{i=1}^n$ *are uniformly bounded in* i *and* n*, then* $\{\mathbb{I}(y_{i,n} > 0)\}_{i=1}^n$ *is a uniformly* L_2*-NED random field on* $\{\epsilon_{i,n}\}$*:* $||\mathbb{I}(y^*_{i,n} > 0) - \mathrm{E}[\mathbb{I}(y^*_{i,n} > 0)|\mathcal{F}_{i,n}(s)]||_2 \leqslant C/s^{(\alpha-d)/3}$ *for some constant* $C > 0$.

Because the joint mixed probability density function of $\{y_{1,n}, \cdots, y_{n,n}\}$ is derived from that of ϵ_n, a log determinant term appears in the log-likelihood function. Taylor's expansion is useful because the log determinant of the Jacobian transformation can be expressed in the form of summation. From there we analyze the NED property for each term in the summation. We summarize some related results in the following two lemmas, one for each setting of the weights matrix.

Lemma 8. *Under Assumption 1, let* $A_n = (a_{ij,n})$ *be an* $n \times n$ *nonstochastic matrix with all elements nonnegative and* $a_{ij,n} = 0$ *when* $d(i,j) > \bar{d}_0 > 0$*, where* $d(i,j)$ *is the distance between individuals* i *and* j*. Suppose* $\sup_n ||A_n||_\infty \leqslant \eta < 1$ *and, the sequence of random variables* $\{v_{i,n}\}_{i=1}^n$ *satisfies* $-1 \leqslant v_{i,n} \leqslant 1$ *and* $||v_{i,n} - \mathrm{E}[v_{i,n}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2 \leqslant C\eta^m$*, where* $\mathcal{F}_{i,n}(m\bar{d}_0) = \sigma(\{\epsilon_{j,n} : d(j,i) \leqslant m\bar{d}_0\})$*, for some positive constant* $C > 0$*, for all positive integers* m*'s,* i*'s and* n*'s. Denote* $G_n = diag(v_{1,n}, \cdots, v_{n,n})$*.*

Then, for any positive integer l*,*

(i) $\{g_{i,ln} \equiv (G_nA_nG_n)^l_{ii}\}_{i=1}^n$ *satisfies* $||g_{i,ln} - \mathrm{E}[g_{i,ln}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2 < (2\eta \max(1, C)/(1-\eta))\eta^m$*;*

(ii) $\{u_{i,n} \equiv [(I_n - G_nA_nG_n)^{-1}G_nA_nG_n]^l_{ii}\}_{i=1}^n$ *satisfies* $||u_{i,n} - \mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2 \leqslant \overline{C_l}m^l\eta^m$ *for some constant* $\overline{C_l} > 0$*.*

Lemma 9. *Under Assumption 1, let* $A_n = (a_{ij,n})$ *be an* $n \times n$ *nonstochastic matrix with* $0 \leqslant a_{ij,n} \leqslant C_0 d(i,j)^{-\alpha}$ *for some positive constants* C_0 *and* α*, where* $d(i,j)$ *is the distance between individuals* i *and* j*. Suppose* $\sup_n ||A_n||_\infty < \eta < 1$ *and, for all positive integer numbers* l*,* $\sup_n ||A_n^l||_1 \leqslant C_2 l \eta^l$ *for some constant* C_2*.* $\{v_{i,n}\}_{i=1}^n$ *satisfies* $-1 \leqslant v_{i,n} \leqslant 1$ *and* $||v_{i,n} - \mathrm{E}[v_{i,n}|\mathcal{F}_{i,n}(s)]||_2 \leqslant C_1 s^{-p}$ *for some positive constant* $p < \alpha$*, all* i*'s and* n*'s, where* $\mathcal{F}_{i,n}(s) = \sigma(\{\epsilon_{j,n} : d(j,i) \leqslant s\})$*. Denote* $G_n = diag(v_{1,n}, \cdots, v_{n,n})$*.*

Then, for any natural number l*,*

(i) $\{g_{i,ln} \equiv (G_nA_nG_n)^l_{ii}\}_{i=1}^n$ *satisfies* $||g_{i,ln} - \mathrm{E}[g_{i,ln}|\mathcal{F}_{i,n}(s)]||_2 \leqslant C_l s^{-p}$ *for some constant* $C_l > 0$*;*

(ii) $\{u_{i,n} \equiv [(I_n - G_nA_nG_n)^{-1}G_nA_nG_n]^l_{ii}\}_{i=1}^n$ *satisfies* $||u_{i,n} - \mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(s)]||_2 \leqslant \overline{C_l} s^{-p}$ *for some constant* $\overline{C_l} > 0$*.*

We apply Lemmas 8 and 9 directly to the model. Let $\widetilde{W_n} = G_n(Y_n)W_nG_n(Y_n)$, where $G_n(Y_n) = diag(\mathbb{I}(y_{1,n} > 0), \cdots, \mathbb{I}(y_{n,n} > 0))$, and $r_{i,n}(\lambda) = ((I_n - \lambda\widetilde{W_n})^{-1}\widetilde{W_n})_{ii}$. With Lemma 7 and $\eta = \zeta$ as $||\lambda W_n||_\infty \leqslant \zeta$, Lemmas 8 and 9 imply the following results with $l = 1$ and 2 in them:

Proposition 2. *(1) Under Assumptions 1, 2, 3(1), 4 and 5, both* $\{r_{i,n}(\lambda_0)\}_{i=1}^n$ *and* $\{[(I_n - \lambda_0\widetilde{W_n})^{-1}\widetilde{W_n}]_{ii}^2\}_{i=1}^n$ *are uniformly and geometrically* L_2*-NED random fields with coefficients* $s(\zeta^{1/\bar{d}_0})^s$ *and* $s^2(\zeta^{1/\bar{d}_0})^s$ *respectively.*

(2) Under Assumptions 1, 2, 3(2), 4 and 5, $\{r_{i,n}(\lambda_0)\}_{i=1}^n$ *and* $\{[(I_n - \lambda_0\widetilde{W_n})^{-1}\widetilde{W_n}]_{ii}^2\}_{i=1}^n$ *are uniformly* L_2*-NED random fields with coefficient* $1/s^{(\alpha-d)/3}$.

From Theorem 1 in Jenish and Prucha (2012), if $\{Z_{i,n},\ i \in D,\ n \geqslant 1\}$ is a uniformly L_1-NED random field, the base $\{\epsilon_{i,n},\ i \in D,\ n \geqslant 1\}$ is i.i.d., and $\sup_{i,n} ||Z_{i,n}||_p < \infty$ for some $p > 1$, then $\frac{1}{n}\sum_{i=1}^n (Z_{i,n} - \mathrm{E}Z_{i,n}) \xrightarrow{L_1} 0$. Since a uniformly L_2-NED random field is also uniformly L_1-NED, the weak LLN also holds. The CLT of NED random fields require more regularity conditions. Assume $Z = \{Z_{i,n},\ i \in D,\ n \geqslant 1\}$ is a zero-mean random field that is L_2-NED on the *i.i.d.* random field $\epsilon = \{\epsilon_{i,n},\ i \in D,\ n \geqslant 1\}$. If Z satisfies the four conditions: (1) it is uniformly $L_{2+\delta}$ integrable for some $\delta > 0$; (2) $\inf_n \frac{1}{n}\sigma_n^2 > 0$, where $\sigma_n^2 = \mathrm{Var}(\sum_{i=1}^n Z_{i,n})$; (3) NED coefficients satisfy $\sum_{r=1}^\infty r^{d-1}\psi(r) < \infty$; and (4) NED scaling factors satisfy $\sup_{n,i\in D} d_{i,n} < \infty$, then $\sigma_n^{-1}\sum_{i=1}^n Z_{i,n} \xrightarrow{d} N(0,1)$. Thus, with the uniform L_p boundedness and uniformly L_2-NED properties established for our model in this section, we may use the LLN and CLT for spatial NED processes to analyze the consistency and asymptotic distribution of the MLE.

4 The MLE and Consistency

As in Amemiya (1973), we will use the method of ML to estimate the true parameters. The Tobit model is established under the distributional specification that the error terms are normally distributed and the parameter space is compact.

Assumption 6. *For each* n, $\epsilon_{i,n}$*'s are i.i.d.* $N(0, \sigma^2)$ *random variables.*

Assumption 7. *The parameter space* Θ *of* $\theta = (\lambda, \beta', \sigma)'$ *is a compact subset of* R^{K+2}*, where* $\sigma > 0$ *is bounded away from 0.*

Qu and Lee (2013a) show that the log-likelihood function of Y_n is

$$\begin{aligned}\ln L_n(\theta) = &\sum_{i=1}^{n} \mathbb{I}(y_{i,n}=0)\ln[1-\Phi(\frac{\lambda}{\sigma}w_{i\cdot,n}Y_n + x_{i,n}\frac{\beta}{\sigma})] - \frac{1}{2}\ln(2\pi\sigma^2)\sum_{i=1}^{n}\mathbb{I}(y_{i,n}>0) \\ &+ \ln|I_{2,n}-\lambda W_{22,n}| - \frac{1}{2}\sum_{i=1}^{n}\mathbb{I}(y_{i,n}>0)(\frac{1}{\sigma}y_{i,n} - \frac{\lambda}{\sigma}w_{i\cdot,n}Y_n - x_{i,n}\frac{\beta}{\sigma})^2 \\ = &\sum_{i=1}^{n}\mathbb{I}(y_{i,n}=0)\ln\Phi(z_{i,n}(\theta)) - \frac{1}{2}\ln(2\pi\sigma^2)\sum_{i=1}^{n}\mathbb{I}(y_{i,n}>0) \\ &+ \ln|I_{2,n}-\lambda W_{22,n}| - \frac{1}{2}\sum_{i=1}^{n}\mathbb{I}(y_{i,n}>0)z_{i,n}^2(\theta),\end{aligned} \tag{4.1}$$

where $W_{22,n}$ is the principal submatrix of W_n corresponding to the strictly positive $y_{i,n}$'s, $I_{2,n}$ is the identity matrix with the same dimension as $W_{22,n}$, and $z_{i,n}(\theta) \equiv (y_{i,n} - \lambda w_{i\cdot,n}Y_n - x_{i,n}\beta)/\sigma$ is defined in Section 3. Note that the dimension of $I_{2,n}$ is stochastic and so are the positions of elements of $W_{22,n}$ in W_n because the number of positive elements in $\{y_{i,n}\}_{i=1}^n$ and their positions for spatial units are random. Maximizing the log-likelihood function, we obtain the MLE $\hat{\theta}$. Recall $\widetilde{W_n} = G_n(Y_n)W_nG_n(Y_n)$. Notice that $trW_{22,n}^l = tr\widetilde{W_n}^l$, and thus

$$\mathrm{E}\ln|I_{2,n}-\lambda W_{22,n}| = -\mathrm{E}\sum_{l=1}^{\infty}(\lambda^l/l)tr(W_{22,n}^{l+1}) = -\mathrm{E}\sum_{l=1}^{\infty}(\lambda^l/l)tr(\widetilde{W_n}^{l+1}) = \mathrm{E}\ln|I_n - \lambda\widetilde{W_n}|.$$

Because all moments of a normal distribution exist, the condition $\sup_n \mathrm{E}|\epsilon_{i,n}|^p < \infty$ in Lemmas 2 - 6 is satisfied for any natural number p. Thus, $\{y_{i,n}\}_{i=1}^n$, $\{y_{i,n}^*\}_{i=1}^n$, $\{z_{i,n}(\theta)\}_{i=1}^n$, $\{z_{i,n}(\theta)^2\}_{i=1}^n$, $\{\ln\Phi(z_{i,n}(\theta))\}_{i=1}^n$, $\{\phi(z_{i,n}(\theta))/\Phi(z_{i,n}(\theta))\}_{i=1}^n$ and $\{\phi(z_{i,n}(\theta))z_{i,n}(\theta)/\Phi(z_{i,n}(\theta))\}_{i=1}^n$ are all uniformly L_p bounded for any natural number p. In the following lemma, we show that the additional boundedness condition for $\{y_{i,n}^*\}_{i=1}^n$ in Lemma 7 is also satisfied. Consequently, $\{\mathbb{I}(y_{i,n}>0)\}_{i=1}^n$ is a uniform L_2-NED random field.

Lemma 10. *Under Assumptions 1, 2, 3 and 6, the essential supremums of densities of* $\{y_{i,n}^*\}_{i=1}^n$ *are uniformly bounded in* i *and* n*.*

Identification is always important for an econometric model. For MLE, identification is equivalent to $P(\ln L_n(\theta_0) \neq \ln L_n(\theta_1)) > 0$ for any $\theta_1 \neq \theta_0$ (Rothenberg, 1971). We summarize a sufficient identification result in the following proposition.

Proposition 3. *Under Assumptions 2 and 6, if $X_n'X_n$ is invertible, $w_{ij,n} \geqslant 0$ and $w_{ii,n} = 0$ for all i and j, and there exists $j \neq j'$ such that $\sum_{i=1}^n w_{ij,n}^2 \neq \sum_{i=1}^n w_{ij',n}^2$, then $\theta_0 = (\lambda_0, \beta_0, \sigma_0^2)$ is identified.*

To show the consistency of the estimator, we need to strengthen the identification information inequality to the limit.

Assumption 8. $\limsup_{n\to\infty}[\mathrm{E}\ln L_n(\theta) - \mathrm{E}\ln L_n(\theta_0)] < 0$ *for any* $\theta \neq \theta_0$.

Now we can state our result about the consistency of MLE. The detailed proof is in the Appendix.

Theorem 1. *Under Assumptions 1 - 8, the MLE of model (2.2) is consistent.*

5 Asymptotic Normality

To discuss the asymptotic normality of the MLE, we present the first derivatives of the log-likelihood function which are arranged in summation form. Recall $z_{i,n}(\theta) \equiv (y_{i,n} - \lambda w_{i\cdot,n}Y_n - x_{i,n}\beta)/\sigma$.

$$\begin{aligned}\frac{\partial \ln L_n(\theta)}{\partial \lambda} =& -\sum_{i=1}^n \mathbb{I}(y_{i,n}=0)\frac{\phi(z_{i,n}(\theta))w_{i\cdot,n}Y_n}{\sigma\Phi(z_{i,n}(\theta))} - tr[(I_{2,n} - \lambda W_{22,n})^{-1}W_{22,n}] \\ &+ \sum_{i=1}^n \mathbb{I}(y_{i,n}>0)\sigma^{-2}(y_{i,n} - \lambda w_{i\cdot,n}Y_n - x_{i.n}\beta)w_{i\cdot,n}Y_n,\end{aligned}$$

$$\frac{\partial \ln L_n(\theta)}{\partial \beta} = -\sum_{i=1}^n \mathbb{I}(y_{i,n}=0)\frac{\phi(z_{i,n}(\theta))x_{i,n}'}{\sigma\Phi(z_{i,n}(\theta))} + \sum_{i=1}^n \mathbb{I}(y_{i,n}>0)\sigma^{-2}(y_{i,n} - \lambda w_{i\cdot,n}Y_n - x_{i.n}\beta)x_{i,n}',$$

$$\begin{aligned}\frac{\partial \ln L_n(\theta)}{\partial \sigma} =& \sum_{i=1}^n \mathbb{I}(y_{i,n}=0)\sigma^{-2}\frac{\phi(z_{i,n}(\theta))(\lambda w_{i\cdot,n}Y_n + x_{i.n}\beta)}{\Phi(z_{i,n}(\theta)} \\ &- \frac{1}{\sigma}\sum_{i=1}^n \mathbb{I}(y_{i,n}>0) + \sum_{i=1}^n \mathbb{I}(y_{i,n}>0)\sigma^{-3}(y_{i,n} - \lambda w_{i\cdot,n}Y_n - x_{i.n}\beta)^2.\end{aligned}$$

Because

$$tr[(I_{2,n} - \lambda W_{22,n})^{-1}W_{22,n}] = tr[(I_n - \lambda\widetilde{W_n})^{-1}\widetilde{W_n}] = \sum_{i=1}^n r_{i,n}(\lambda),$$

the score can be written in terms of a summation as $\partial \ln L_n(\theta)/\partial\theta = \sum_{i=1}^n q_{i,n}(\theta)$, where

$$q_{i,n}(\theta)=\begin{pmatrix} -\mathbb{I}(y_{i,n}=0)\frac{\phi(z_{i,n}(\theta))w_{i\cdot,n}Y_n}{\sigma\Phi(z_{i,n}(\theta))}+\mathbb{I}(y_{i,n}>0)\sigma^{-1}z_{i,n}(\theta)w_{i\cdot,n}Y_n-r_{i,n}(\lambda) \\ -\mathbb{I}(y_{i,n}=0)\frac{\phi(z_{i,n}(\theta))x'_{i,n}}{\sigma\Phi(z_{i,n}(\theta))}+\mathbb{I}(y_{i,n}>0)\sigma^{-1}z_{i,n}(\theta)x'_{i,n} \\ \mathbb{I}(y_{i,n}=0)\sigma^{-2}\frac{\phi(z_{i,n}(\theta))(\lambda w_{i\cdot,n}Y_n+x_{i.n}\beta)}{\Phi(z_{i,n}(\theta))}-\frac{1}{\sigma}\mathbb{I}(y_{i,n}>0)+\mathbb{I}(y_{i,n}>0)\sigma^{-1}z_{i,n}(\theta)^2 \end{pmatrix}. \tag{5.1}$$

To obtain the asymptotic normality of the MLE, some additional regular conditions are needed.

Assumption 9. *θ_0 is in the interior of Θ.*

Assumption 10. *$\Sigma_0=\lim_{n\to\infty}\Sigma_n$ exists and is nonsingular, where $\Sigma_n=\frac{1}{n}\mathrm{Var}\sum_{i=1}^n q_{i,n}(\theta_0)$.*

Assumption 11. *$0\leqslant w_{ij,n}\leqslant C_0/d(i,j)^\alpha$, where $\alpha>7d$.*

Recall that under the two different settings in Assumption 3, we have different NED coefficients. Under Assumption 3(1), $\{q_{i,n}(\theta_0)\}_{i=1}^n$ is a uniformly and geometrically L_2-NED random field, so Assumption 11 is not needed. Under Assumption 3(2)(ii), Assumption 11 is needed to derive the asymptotic distribution. From the expression of $q_{i,n}(\theta)$, most terms are products of two NED random fields. Though the product of NED random fields remains an NED random field, the NED coefficient usually decreases slower. Thus the NED coefficient of $q_{i,n}(\theta)$ is slower than the order $O(1/s^{(\alpha-d)/3})$ of $\mathbb{I}(y_{i,n}=0)$. Actually, we show that the NED coefficient of the Euclidean norm of $q_{i,n}(\theta)$ has the order $(1/s^{\alpha-d})^{(1-\gamma)/6}$, for any $0<\gamma<1$. To satisfy the condition $\sum_{s=1}^\infty s^{d-1}(1/s^{\alpha-d})^{(1-\gamma)/6}<\infty$, we thus assume $\alpha>7d$.

Lemma 11. *In addition to Assumptions 1, 2, 4 and 6 -10, suppose either 3(1), or Assumptions 3(2) and 11 hold, then $\frac{1}{\sqrt{n}}\sum_{i=1}^n q_{i,n}(\theta_0)\xrightarrow{d}N(0,\Sigma_0)$.*

Theorem 2. *Under Assumptions 1, 2, 4 and 6 -10, if either Assumption 3(1), or Assumptions 3(2) and 11 hold, then the MLE of model (2.2) has $\sqrt{n}(\hat{\theta}-\theta_0)\xrightarrow{d}N(0,\Sigma_0^{-1})$.*

6 Monte Carlo Simulations

In this section, we perform simulations to study the finite sample performance of the MLE and test its robustness. In the experiments, we consider the equation $Y_n=F(\lambda W_nY_n+\beta_1+\beta_2X_n+\epsilon_n)$, where X_n is a column vector. We try two groups of true parameters with different values of

λ: $(\lambda_0, \beta_{10}, \beta_{20}, \sigma_0) = (0.3, -1, 2, 2)$ and $(0.5, -1, 2, 2)$ to generate the data. Elements of X_n are independently generated from $N(0.2, 0.25)$. The error terms are i.i.d. $N(0, \sigma_0^2)$ distributed.

Here we discuss the generation of the weights matrix W_n. The connection relationship of the 3142 counties in the U.S. can be found in "Contiguous County File, 1991: [United States]". Thus we have a 3142×3142 matrix W_0 whose elements are one if the corresponding counties are contiguous; otherwise, zero. When we generate a uniform random natural number i between 1 and 2143, then we use the entries of W_0 that are between the i^{th} and the $(i+999)^{th}$ row and between the i^{th} and the $(i+999)^{th}$ column to form a 1000×1000 submatrix. Finally, we row normalize it to obtain W_n. Similarly we obtain W_n when the sample sizes are 100, 200 and 500, except that W_0 now is generated from 760 counties in the 10 Upper Great Plains States[8], rather than all the counties in the U.S. To do so, we can have more nonzero elements when the sample size is smaller. With the data of W_n, X_n and ϵ_n, we generate the data of Y_n by contraction mapping. The iteration stops when $||Y_n - F(\lambda W_n Y_n + \beta_1 + \beta_2 X_n + \epsilon_n)||_\infty < 10^{-6}$.

We can obtain the mean and standard deviation of the estimators based on 1000 replications for each of the experiments.

Table 1: Estimation Results When the Errors are Normally Distributed

	λ	β_1	β_2	σ	λ	β_1	β_2	σ
true	0.3	-1	2	2	0.5	-1	2	2
$n = 100$	0.2267	-0.9524	2.0184	1.9452	0.4275	-0.9495	2.0176	1.9509
	(0.2758)	(0.3870)	(0.5745)	(0.2421)	(0.2345)	(0.3658)	(0.5587)	(0.2324)
$n = 200$	0.2651	-0.9671	1.9978	1.9762	0.4663	-0.9667	1.9998	1.9791
	(0.1903)	(0.2808)	(0.3406)	(0.1697)	(0.1580)	(0.2633)	(0.3295)	(0.1648)
$n = 500$	0.2805	-0.9833	2.0087	1.9920	0.4810	-0.9821	2.0097	1.9926
	(0.1311)	(0.1789)	(0.2331)	(0.1047)	(0.1101)	(0.1690)	(0.2243)	(0.0993)
$n = 1000$	0.2914	-0.9853	1.9927	1.9944	0.4912	-0.9869	1.9954	1.9963
	(0.0925)	(0.1273)	(0.1554)	(0.0716)	(0.0767)	(0.1210)	(0.1520)	(0.0693)

[8]The ten states include Colorado, Iowa, Kansas, Minnesota, Missouri, Montana, Nebraska, North Dakota, South Dakota, and Wyoming

As can be seen in Table 1, as the sample size n increases, both the bias and the standard error of the estimator decrease. The simulation results verify the consistency of the MLE.

To test the robustness of the MLE, we also try error terms with different distributions. For the purpose of comparison, we take $(\lambda_0, \beta_{10}, \beta_{20}, \sigma_0) = (0.5, -1, 2, 2)$, which is the same as the second true parameter vector in Table 1. We try four different distributions: uniform distribution $U(-2\sqrt{3}, 2\sqrt{3})$, $2\sqrt{0.6}$ times student $t(5)$ distribution, a mixed normal distribution (with half probability $N(8/\sqrt{17}, 4/17)$, half probability $N(-8/\sqrt{17}, 4/17)$), and $4\sqrt{2}\mathrm{B}(0.5, 0.5) - 2\sqrt{2}$, where $\mathrm{B}(0.5, 0.5)$ is a two-parameter Beta distribution. The other settings are the same as the previous one. Simple calculations show that all four of these distributions have standard deviations $\sigma_0 = 2$. Notice that the densities of the last two distributions are not single peak. The mixed normal distribution has two peaks, while $\mathrm{B}(0.5, 0.5)$ has a U-shaped density.

Table 2: Estimation When Error Terms Are Not Normally Distributed

distribution	$U(-2\sqrt{3}, 2\sqrt{3})$				$2\sqrt{0.6}t(5)$			
	λ	β_1	β_2	σ	λ	β_1	β_2	σ
$n = 100$	0.4024 (0.2260)	-0.7500 (0.3423)	1.8077 (0.5074)	1.9155 (0.1700)	0.4630 (0.2187)	-1.1080 (0.3914)	2.1650 (0.5631)	1.9451 (0.3350)
$n = 200$	0.4431 (0.1545)	-0.7974 (0.2501)	1.8298 (0.3190)	1.9541 (0.1190)	0.4778 (0.1645)	-1.1099 (0.2898)	2.1390 (0.3659)	1.9944 (0.2690)
$n = 500$	0.4594 (0.1160)	-0.7987 (0.1656)	1.8010 (0.2078)	1.9483 (0.0745)	0.5001 (0.1197)	-1.1471 (0.1937)	2.1597 (0.2497)	2.0265 (0.1696)
$n = 1000$	0.4650 (0.0778)	-0.8111 (0.1076)	1.8207 (0.1415)	1.9570 (0.0514)	0.5178 (0.0761)	-1.1685 (0.1403)	2.1693 (0.1692)	2.0339 (0.1373)
distribution	mixed normal distribution				$4\sqrt{2}\mathrm{B}(0.5, 0.5) - 2\sqrt{2}$			
	λ	β_1	β_2	σ	λ	β_1	β_2	σ
$n = 100$	0.3793 (0.2360)	-0.5138 (0.3151)	1.5092 (0.4688)	1.8560 (0.1446)	0.4002 (0.2305)	-0.6649 (0.3380)	1.6847 (0.4832)	1.8926 (0.1518)
$n = 200$	0.4144 (0.1569)	-0.6093 (0.2393)	1.5864 (0.2995)	1.9470 (0.0997)	0.4206 (0.1702)	-0.7166 (0.2481)	1.7343 (0.3165)	1.9508 (0.1141)
$n = 500$	0.4306 (0.1170)	-0.5813 (0.1561)	1.5285 (0.1950)	1.9078 (0.0608)	0.4448 (0.1138)	-0.7182 (0.1600)	1.7246 (0.2046)	1.9334 (0.0660)
$n = 1000$	0.4372 (0.0783)	-0.6027 (0.1071)	1.5788 (0.1324)	1.9163 (0.0438)	0.4540 (0.0774)	-0.7290 (0.1109)	1.7276 (0.1431)	1.9387 (0.0468)

$(\lambda_0, \beta_{10}, \beta_{20}, \sigma_0) = (0.5, -1, 2, 2)$.
mixed normal distribution: half probability $N(8/\sqrt{17}, 4/17)$, half probability $N(-8/\sqrt{17}, 4/17)$.

The estimation results are summarized in Table 2. When the distributions are uniform, mixed normal or Beta, the bias of the estimator is much larger than that under the normal distribution.

When the error is $t(5)$ distributed, the bias is less than that under the uniform, mixed normal or Beta distributions, but it is obviously larger than that under the normal distribution. This simple observation indicates that the MLE can be consistent only when the distribution is correctly specified as normal.

7 Conclusion

This paper examines the ML estimation of a spatial autoregressive Tobit model. We establish the NED properties of the dependent variable and some other random fields. With the LLN and CLT of the NED random fields, we establish the consistency and the asymptotic normality of the MLE of this model. Monte Carlo simulations verify the consistency of MLE. The simulations also show that if the shapes of the distribution of errors are rather different from that of the normal distribution, MLE may cause large bias. With the consistency and asymptotic normality of MLE, we can use ML to estimate SART models when we study empirical problems involving spatial correlation and censored data.

While focusing on the large sample properties of the MLE of the SART model, we realize that there are a few limitations on the study. Most notably, we only consider the case in which the error terms are independently, identically and normally distributed. Nevertheless, this paper develops solid foundations for future extensions of this paper: (1) estimation of the model with unknown form of heteroskedasticity; (2) estimation of the model when the error terms are distributional free; and (3) estimation of other types of Tobit models (see Amemiya, 1985), such as sample selection models with interactions, in the framework of SAR models.

Appendices

A Some Useful Lemmas

Lemma A.1. *(Lemma 17.15 in Davidson (1994)) Let B and ρ be two non-negative random variables and assume $||\rho||_q < \infty$, $||B||_p < \infty$, and $||B\rho||_r < \infty$, for $q^{-1} + p^{-1} = 1$, $q \geqslant 1$ and $r > 2$. Then*

$$||B\rho||_2 \leqslant 2(||\rho||_q^{r-2}||B||_p^{r-2}||B\rho||_r^r)^{1/(2r-2)}.$$

Lemma A.2. *(Generalization of Corollary 4.3 (b), Gallant and White (1988)) If, for all i and n, $||Y_{i,n}||_{2r} \leqslant \Delta < \infty$ and $||Z_{i,n}||_{2r} \leqslant \Delta < \infty$ for some $r > 2$, $||Y_{i,n} - \mathrm{E}[Y_{i,n}|\mathcal{F}_{i,n}(s)]||_2 \leqslant d_{i,Yn}\psi(s)$ and $||Z_{i,n} - \mathrm{E}[Z_{i,n}|\mathcal{F}_{i,n}(s)]||_2 \leqslant d_{i,Zn}\psi(s)$, then*

$$||Y_{i,n}Z_{i,n} - \mathrm{E}[Y_{i,n}Z_{i,n}|\mathcal{F}_{i,n}(s)]||_2 \leqslant d_{i,n}\tilde{\psi}(s),$$

where $d_{i,n} = 2^{(3r-2)/(r-1)}(d_{i,Zn} + d_{i,Yn})^{(r-2)/(2r-2)}\Delta^{(3r-2)/(2r-2)}$ and $\tilde{\psi}(s) = \psi(s)^{(r-2)/(2r-2)}$.

Specifically, if $\{Y_{i,n}\}$ and $\{Z_{i,n}\}$ are both uniformly L_{2r} bounded, and uniformly and geometrically L_2-NED, then $\{Y_{i,n}Z_{i,n}\}$ is uniformly and geometrically L_2-NED.

Lemma A.3. *If $\{\epsilon_{i,n}\}_{i=1}^n$'s are identically and independently distributed with $\mathrm{E}|\epsilon_{i,n}|^p < \infty$, $M = (m_{ij,n})$ is a nonstochastic $n \times n$ matrix with $||M||_\infty \leqslant B$, then $\mathrm{E}[(\sum_{j=1}^n |m_{ij,n}\epsilon_{j,n}|)^p] \leqslant \mathrm{E}|\epsilon_{1,n}|^p B^p$.*

Proof of Lemma A.3: Because $\prod_{j=1}^n \mathrm{E}|\epsilon_{j,n}|^{l_j} \leqslant \mathrm{E}|\epsilon_{1,n}|^{l_1+\cdots+l_n}$ when all l_i's are nonnegative,

$$\begin{aligned}
&\mathrm{E}[(\sum_{j=1}^n |m_{ij,n}\epsilon_{j,n}|)^p] = \mathrm{E}\sum_{l_1+\cdots+l_n=p}\prod_{j=1}^n |m_{ij,n}\epsilon_{j,n}|^{l_j} \\
=& \sum_{l_1+\cdots+l_n=p}\prod_{j=1}^n |m_{ij,n}|^{l_j}\mathrm{E}|\epsilon_{j,n}|^{l_j} \leqslant \sum_{l_1+\cdots+l_n=p}\left(\prod_{j=1}^n |m_{ij,n}|^{l_j}\right)\mathrm{E}|\epsilon_{1,n}|^{l_1+\cdots+l_n} \\
=&\mathrm{E}|\epsilon_{1,n}|^p(\sum_{j=1}^n |m_{ij,n}|)^p \leqslant \mathrm{E}|\epsilon_{1,n}|^p B^p.
\end{aligned}$$

□

Lemma A.4. *Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $n \times n$ matrices, and let e be a column vector of dimension n. If $|A|_{\max} \equiv \max_{i,j} |a_{ij}|$, then for any positive integer l, $||[(A+B)^l - B^l]e||_\infty \leqslant |A|_{\max} \sum_{h=0}^{l-1} ||B||_\infty^h ||A+B||_1^{l-1-h} ||e||_1$.*

Proof of Lemma A.4: Let $W = A + B$. By expansion, $W^l - B^l = \sum_{h=0}^{l-1} B^h A W^{l-1-h}$. For any matrix M of dimension n, it is simple to see that $||Me||_\infty \leqslant |M|_{\max} ||e||_1$. Thus, for any integer $h = 0, \cdots, l-1$,

$$\begin{aligned} &||B^h A W^{l-1-h} e||_\infty \leqslant ||B^h||_\infty ||A W^{l-1-h} e||_\infty \\ \leqslant &||B^h||_\infty |A|_{\max} ||W^{l-1-h} e||_1 = |A|_{\max} ||B^h||_\infty ||W||_1^{l-1-h} ||e||_1. \end{aligned}$$

Together, we have the result. □

Lemma A.5. *$G(x)$: Domain($\subset R$) $\to R$ satisfies $|G(x_1) - G(x_2)| \leqslant C_1(|x_1|^a + |x_2|^a + 1)|x_1 - x_2|$ for some integer $a \geqslant 1$. If $\{u_{i,n}\}_{i=1}^n$ is a random field with $||u_{i,n} - \mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(s)]||_2 \leqslant C_2 \psi(s)$ for all i and n, and $\sup_{i,n} ||u_{i,n}||_p < \infty$ for some $p > 2a + 2$. Then $||G(u_{i,n}) - \mathrm{E}[G(u_{i,n})|\mathcal{F}_{i,n}(s)]||_2 \leqslant C\psi(s)^{(p-2a-2)/(2p-2a-2)}$.*

Proof of Lemma A.5: Because $p/(a+1) > 2$, with Lemma A.1, we have

$$\begin{aligned} &||G(u_{i,n}) - \mathrm{E}[G(u_{i,n})|\mathcal{F}_{i,n}(s)]||_2 \leqslant ||G(u_{i,n}) - G(\mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(s)])||_2 \\ \leqslant &C_1 ||(|u_{i,n}|^a + |\mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(s)]|^a + 1) \cdot (u_{i,n} - \mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(s)])||_2 \\ \leqslant &2C_1 ||B_{i,n}||_2^{(p-2a-2)/(2p-2a-2)} ||\rho_{i,n}||_2^{(p-2a-2)/(2p-2a-2)} ||B_{i,n}\rho_{i,n}||_{p/(a+1)}^{p/(2p-2a-2)}, \end{aligned}$$

where $B_{i,n} = |u_{i,n}|^a + |\mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(s)]|^a + 1$ and $\rho_{i,n} = u_{i,n} - \mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(s)]$. By Jensen's inequality, $||B_{i,n}||_{p/a} \leqslant 2||u_{i,n}||_p^a + 1$ and $||\rho_{i,n}||_p \leqslant 2||u_{i,n}||_p$ for $p \geqslant 1$.[9] By generalized Hölder's inequality, we have $\sup_{i,n} ||B_{i,n}\rho_{i,n}||_{p/(a+1)} \leqslant \sup_{i,n} ||B||_{p/a} ||\rho_{i,n}||_p \leqslant \sup_{i,n} (2||u_{i,n}||_p^a + 1) \cdot 2||u_{i,n}||_p < \infty$. Thus the conclusion holds. □

Lemma A.6. *Assume $u(v) : R^p \to R^p$ satisfies $||u(v_1) - u(v_2)|| \leqslant C||v_1 - v_2||$ for some constant $C > 0$ and for all v_1 and v_2.*

(1) If $\{f_n(u) : R^p \to R\}_{i=1}^n$ is equicontinuous with respect to u, then $\{f_n(u(v)) : R^p \to R\}_{i=1}^n$ is equicontinuous with respect to v.

[9] By Jensen's inequality, for $a \geqslant 1$ and $b \geqslant 1$, $||E^a(|x||\mathcal{F})||_b \leqslant ||x||_{ab}^a$.

(2) If $\{f_n(u) : R^p \to R\}_{i=1}^n$ is stochastically equicontinuous with respect to u, then $\{f_n(u(v)) : R^p \to R\}_{i=1}^n$ is stochastically equicontinuous with respect to v.

Proof: (1) Because $\{f_n(u) : R^p \to R\}_{i=1}^n$ is equicontinuous, for any constant $\epsilon > 0$, there exists $\delta > 0$ such that when $||u_1 - u_2|| < \delta$, we have $|f_n(u_1) - f_n(u_2)| < \epsilon$ for any n. When $||v_1 - v_2|| < \delta/C$, we have $||u(v_1) - u(v_2)|| \leqslant \delta$, thus $|f_n(u(v_1)) - f_n(u(v_2))| < \epsilon$ for any n. (2) Because $\{f_n(u) : R^p \to R\}_{i=1}^n$ is stochastically equicontinuous, for any constant $\epsilon > 0$, there exists $\delta > 0$ such that $\limsup_{n\to\infty} P(\sup_{||u_1-u_2||<\delta} |f_n(u_1) - f_n(u_2)| > \epsilon) < \epsilon$. When $||v_1 - v_2|| < \delta/C$, we have $||u(v_1) - u(v_2)|| \leqslant \delta$, thus $\limsup_{n\to\infty} P(\sup_{||v_1-v_2||<\delta/C} |f_n(u_1) - f_n(u_2)| > \epsilon) < \epsilon$. □

Lemma A.7. *Assume $f : D(\subset R) \to R$ satisfies $|f(x_1) - f(x_2)| \leqslant C(|x_1|^a + |x_2|^a + 1)|x_1 - x_2|$ for some constants $a \geqslant 1$, $C > 0$, and for all $x_1,\ x_2 \in D$. If the random field $\{x_{i,n}\}_{i=1}^n \subset R^K$ satisfies $\sup_{i,j,n} ||x_{ij,n}||_{\max(2a,4)} < \infty$ and $\sup_{i,j,n} ||h(x_{i,n})||_4 < \infty$, then $\{\frac{1}{n}\sum_{i=1}^n f(x_{i,n}\theta)h(x_{i,n})\}_{n=1}^\infty$ is stochastically equicontinuous with respect to θ, where the parameter space $\Theta(\subset R^K)$ of θ is bounded.*

Proof:

$$
\begin{aligned}
&|\frac{1}{n}\sum_{i=1}^n f(x_{i,n}\theta_1)h(x_{i,n}) - \frac{1}{n}\sum_{i=1}^n f(x_{i,n}\theta_2)h(x_{i,n})| \\
\leqslant&\frac{1}{n}\sum_{i=1}^n |[f(x_{i,n}\theta_1) - f(x_{i,n}\theta_2)]h(x_{i,n})| \\
\leqslant&\frac{C}{n}\sum_{i=1}^n (|x_{i,n}\theta_1|^a + |x_{i,n}\theta_2|^a + 1)|x_{i,n}(\theta_1 - \theta_2)| \cdot |h(x_{i,n})| \\
\leqslant&\frac{C}{n}\sum_{i=1}^n \{K^{a-1}\sum_{j=1}^K |x_{ij,n}\theta_{1j}|^a + K^{a-1}\sum_{j=1}^K |x_{ij,n}\theta_{2j}|^a + 1\}|h(x_{i,n})| \cdot (\sum_{k=1}^K |x_{ik,n}| \cdot |\theta_{1k} - \theta_{2k}|) \\
\leqslant&\frac{C}{n}\sum_{i=1}^n \{2K^{a-1}\sum_{j=1}^K |x_{ij,n}|^a(\sup_{\theta\in\Theta}|\theta_j|)^a + 1\}|h(x_{i,n})| \cdot (\sum_{k=1}^K |x_{ik,n}| \cdot |\theta_{1k} - \theta_{2k}|),
\end{aligned}
$$

where the third inequality comes from C_r-inequality. By Cauchy's inequality[10], we have

$$
\begin{aligned}
&||\frac{1}{n}\sum_{i=1}^{n}\{2K^{a-1}\sum_{j=1}^{K}|x_{ij,n}|^a(\sup_{\theta\in\Theta}|\theta_j|)^a+1\}|h(x_{i,n})|\cdot|x_{ik,n}|||_1\\
\leqslant&\frac{1}{n}\sum_{i=1}^{n}||2K^{a-1}\sum_{j=1}^{K}|x_{ij,n}|^a(\sup_{\theta\in\Theta}|\theta_j|)^a+1||_2\cdot||h(x_{i,n})x_{ik,n}||_2\\
\leqslant&\frac{1}{n}\sum_{i=1}^{n}||2K^{a-1}\sum_{j=1}^{K}|x_{ij,n}|^a(\sup_{\theta\in\Theta}|\theta_j|)^a+1||_2\cdot||h(x_{i,n})||_4\cdot||x_{ik,n}||_4<\infty.
\end{aligned}
$$

Hence, by the Markov inequality, $\frac{1}{n}\sum_{i=1}^{n}\{2K^{a-1}\sum_{j=1}^{K}|x_{ij,n}|^a(\sup_{\theta\in\Theta}|\theta_j|)^a+1\}|h(x_{i,n})\cdot|x_{ik,n}| = O_p(1)$, which does not depend on θ. Consequently, $\{\frac{1}{n}\sum_{i=1}^{n}f(x_{i,n}\theta)h(x_{i,n})\}_{n=1}^{\infty}$ is stochastically equicontinuous with respect to θ by Lemma 1 (a) in Andrews (1992). □

B Proofs

In this paper, we will use C, C_1, $C_2\cdots$ to represent some positive constants, which can be different in different places.

B.1 The Proof for Section 2

Proof of Lemma 1: By Lemma A.1 in Jenish and Prucha (2009), $|\{j : m\leqslant d(i,j)<m+1\}|\leqslant Cm^{d-1}$ for some constant $C>0$ when $m\geqslant 1$. Then $\Gamma=\lambda_m\sup_n||W_n||_1<\infty$ comes from

$$
\begin{aligned}
&\sup_n||W_n||_1=\sup_{n,j}\sum_{i=1}^{n}w_{ij,n}=\sup_{n,j}\sum_{m=1}^{\infty}\sum_{i:m\leqslant d(i,j)<m+1}w_{ij,n}\\
=&\sup_j\sum_{m=1}^{\infty}\sum_{i:m\leqslant d(i,j)<m+1}C_0m^{-\alpha}\leqslant\sum_{m=1}^{\infty}Cm^{d-1}C_0m^{-\alpha}<\infty.
\end{aligned}
$$

$||W_n^l||_1=\max_{i=1,\cdots n,}||W_n^le_i||_1$, where $e_i=(0,\cdots,0,1,0,\cdots 0)'$ is the ith unit vector of dimension n. Let $\iota_n=\sum_{i=1}^{n}e_i$. Thus $||W_n^le_i||_1=\iota_n'W_n^le_i$, since all elements of W_n are non-negative. Without loss of generality, we assume that only the first N columns might satisfy $\Gamma\geqslant\lambda_m\sum_i w_{ij,n}>\zeta$ and the remaining $n-N$ columns must satisfy $\lambda_m\sum_i w_{ij,n}\leqslant\zeta$. Then,

[10]If random variables X and Y satisfy with $||X||_{2p}<\infty$ and $||Y||_{2p}<\infty$ for $p\geqslant 1$, then $||XY||_p\leqslant||X||_{2p}||Y||_{2p}$.

for $l \geqslant 1$,

$$
\begin{aligned}
&\iota_n' W_n^l e_i = (\iota_n' w_{\cdot 1,n}, \cdots, \iota_n' w_{\cdot n,n}) W_n^{l-1} e_i \\
=&(\iota_n' w_{\cdot 1,n}, \cdots, \iota_n' w_{\cdot N,n}, 0, \cdots, 0) W_n^{l-1} e_i + (0, \cdots, 0, \iota_n' w_{\cdot (N+1),n}, \cdots, \iota_n' w_{\cdot n,n}) W_n^{l-1} e_i \\
\leqslant&||(\iota_n' w_{\cdot 1,n}, \cdots, \iota_n' w_{\cdot N,n}, 0, \cdots, 0)'||_1 ||W_n^{l-1} e_i||_\infty \\
&+ ||(0, \cdots, 0, \iota_n' w_{\cdot (N+1),n}, \cdots, \iota_n' w_{\cdot n,n})'||_\infty ||W_n^{l-1} e_i||_1 \\
\leqslant&N \lambda_m^{-1} \Gamma ||W_n^{l-1} e_i||_\infty + \lambda_m^{-1} \zeta ||W_n^{l-1} e_i||_1 \\
\leqslant&N \lambda_m^{-1} \Gamma ||W_n^{l-1}||_\infty ||e_i||_\infty + \lambda_m^{-1} \zeta ||W_n^{l-1} e_i||_1 \\
\leqslant&N \lambda_m^{-l} \Gamma \zeta^{l-1} + \lambda_m^{-1} \zeta ||W_n^{l-1} e_i||_1.
\end{aligned}
$$

Thus $||W_n^l e_i||_1 \leqslant N \lambda_m^{-l} \Gamma \zeta^{l-1} + \lambda_m^{-1} \zeta ||W_n^{l-1} e_i||_1$. With recursion and $||W_n e_i||_1 \leqslant \lambda_m^{-1} \Gamma$, we obtain that $\lambda_m^l ||W_n^l e_i||_1 \leqslant [(l-1)N + 1] \Gamma \zeta^{l-1} \leqslant lN \Gamma \zeta^{l-1}$. □

B.2 The Proof for Section 3

Proof of Lemma 2: Even though $F(x) = \max(0, x)$ in Eq. (2.1) is not differentiable, we can apply the mean value theorem of a convex function (see Wegge (1974)) to this $F(\cdot)$. Consider $Y_1 = F(\lambda W_n Y_1 + \eta_1)$ and $Y_2 = F(\lambda W_n Y_2 + \eta_2)$. Then $Y_1 - Y_2 = F(\lambda W_n Y_1 + \eta_1) - F(\lambda W_n Y_2 + \eta_2) = f_{D_n} \cdot [\lambda W_n (Y_1 - Y_2) + (\eta_1 - \eta_2)]$, where $f_{D_n} = diag(f_{i,n})$ is a diagonal matrix whose i-th diagonal element is some subgradient of $F(\cdot)$ at some points between $\lambda w_{i\cdot,n} y_{i,1} + \eta_{i,1}$ and $\lambda w_{i\cdot,n} y_{i,2} + \eta_{i,2}$. The subgradient of $F(\cdot)$ is always between 0 and 1. Then, $(I_n - \lambda f_{D_n} W_n)$ is invertible because $||\lambda f_{D_n} W_n||_\infty < 1$ under Assumption 2, and hence,

$$
\begin{aligned}
&||Y_1 - Y_2||_\infty = ||(I_n - \lambda f_{D_n} W_n)^{-1} f_{D_n} (\eta_1 - \eta_2)||_\infty \\
\leqslant&||(I_n - \lambda f_{D_n} W_n)^{-1} f_{D_n}||_\infty ||\eta_1 - \eta_2||_\infty \\
=&||\sum_{l=0}^{\infty} \lambda^l f_{D_n}^l W_n^l f_{D_n}||_\infty ||\eta_1 - \eta_2||_\infty \\
\leqslant&||\sum_{l=0}^{\infty} |\lambda|^l f_{D_n}^l W_n^l f_{D_n}||_\infty ||\eta_1 - \eta_2||_\infty \\
\leqslant&||\sum_{l=0}^{\infty} \lambda_m^l W_n^l||_\infty ||\eta_1 - \eta_2||_\infty,
\end{aligned}
\tag{B.1}
$$

where the third inequality comes from $|\lambda|^l f_{D_n}^l W_n^l f_{D_n} \leqslant^* \lambda_m^l W_n^l$ and $(a_{ij}) \leqslant^* (b_{ij})$ means $a_{ij} \leqslant b_{ij}$ for all i's and j's. Define $M_n = (m_{ij,n}) \equiv \sum_{l=0}^{\infty} \lambda_m^l W_n^l = (I_n - \lambda_m W_n)^{-1}$. Then $||M_n||_\infty \leqslant \sum_{l=0}^{\infty} ||\lambda_m W_n||_\infty^l \leqslant \sum_{l=0}^{\infty} \zeta^l = 1/(1-\zeta)$. Thus, Y is a Lipschitz function of η. Denote the solution of $Y_n = F(\lambda_0 W_n Y_n + X_n\beta_0 + \epsilon_n)$ as $Y_n(\epsilon_n)$. We note that $Y = F(\lambda_0 W_n Y)$ implies that $Y = 0$ as the solution. Because Y_n is a Lipschitz function of η_n, where $\eta_n = X_n\beta_0 + \epsilon_n$, and $\{x_{i,n}\}_{i=1}^n$ is uniformly bounded, $\{y_{i,n}(0)\}_{i=1}^n$ is also uniformly bounded as $||Y_n(0)||_\infty \leqslant \frac{1}{1-\zeta}||X_n\beta_0||_\infty$. The relation $Y_1 - Y_2 = \sum_{l=0} \lambda^l f_{D_n}^l W_n^l f_{D_n}(\eta_1 - \eta_2)$ implies that $|y_{i,n}(\epsilon_n) - y_{i,n}(0)| \leqslant \sum_{j=1}^n |m_{ij,n}\epsilon_{i,n}|$. From Lemma A.3, $\mathrm{E}[(\sum_{j=1}^n |m_{ij,n}\epsilon_{j,n}|)^p] \leqslant \mathrm{E}|\epsilon_{1,n}|^p/(1-\zeta)^p$. Because $|y_{i,n}(\epsilon_n) - y_{i,n}(0)| \leqslant \sum_{j=1}^n |m_{ij,n}\epsilon_{j,n}|$, it follows from C_r-inequality that

$$
\begin{aligned}
&\sup_{i,n} \mathrm{E}[|y_{i,n}(\epsilon_n)|^p] \leqslant \sup_{i,n} \mathrm{E}[(|y_{i,n}(0)| + \sum_{j=1}^n |m_{ij,n}\epsilon_{j,n}|)^p] \\
\leqslant &\sup_{i,n} 2^{p-1}\{|y_{i,n}(0)|^p + \mathrm{E}[(\sum_{j=1}^n |m_{ij,n}\epsilon_{j,n}|)^p]\} \\
\leqslant &\sup_{i,n} 2^{p-1}[|y_{i,n}(0)|^p + \mathrm{E}|\epsilon_{1,n}|^p/(1-\zeta)^p] < \infty.
\end{aligned}
$$

□

Proof of Proposition 1: Define $H(Y) = F(\lambda_0 W_n Y + X_n\beta_0 + \epsilon_n)$.

(1) Because $|F(a) - F(b)| \leqslant |a - b|$, $|H_{j,n}(Y_1) - H_{j,n}(Y_2)| \leqslant \lambda_m |\sum_{k=1}^n w_{jk,n}(Y_{k,1} - Y_{k,2})|$. It follows that

$$|H_{j,n}(Y_1) - H_{j,n}(Y_2)|^2 \leqslant \lambda_m^2 \sum_{k=1}^n \sum_{l=1}^n w_{jk,n} w_{jl,n} |(Y_{k,1} - Y_{k,2})(Y_{l,1} - Y_{l,2})|$$

and

$$
\begin{aligned}
&\mathrm{E}(|H_{j,n}(Y_1) - H_{j,n}(Y_2)|^2) \\
\leqslant &\lambda_m^2 \sum_{k=1}^n \sum_{l=1}^n w_{jk,n} w_{jl,n} ||Y_{k,1} - Y_{k,2}||_2 \cdot ||Y_{l,1} - Y_{l,2}||_2 \qquad \text{(B.2)} \\
\leqslant &\lambda_m^2 (\sum_{k=1}^n w_{jk,n})^2 \max_{k=1,\cdots,n} ||Y_{k,1} - Y_{k,2}||_2^2.
\end{aligned}
$$

Hence, $||H_{j,n}(Y_1) - H_{j,n}(Y_2)||_2 \leqslant \lambda_m ||W_n||_\infty \max_{k=1,\cdots,n} ||Y_{k,1} - Y_{k,2}||_2 \leqslant \zeta \max_{k=1,\cdots,n} ||Y_{k,1} -$

$Y_{k,2}||_2$. The contraction mapping H_n has the unique fixed point $Y_n = H_n(Y_n)$. Take the initial value $Y_n^{(0)} = (F(x_1\beta), \cdots, F(x_n\beta))'$. By iteration, $Y_n^{(m+1)} = H_n(Y_n^{(m)})$. Because the distance of direct interaction is no more than $\bar{d}_0$, $Y_{i,n}^{(m)}$ will be functions of $\epsilon_{j,n}$'s with units j's located within the distance $m\bar{d}_0$ from i, for each $m = 1, 2, \cdots$. Thus, we have

$$||y_{j,n} - \mathrm{E}(y_{j,n}|\mathcal{F}_{j,n}(m\bar{d}_0))||_2 \leqslant ||y_{j,n} - Y_{j,n}^{(m)}||_2 \leqslant \zeta \max_{k=1,\cdots,n} ||y_{k,n} - Y_{k,n}^{(m-1)}||_2 \leqslant c_n \zeta^m,$$

where $c_n = \max_{k=1,\cdots,n} ||y_{k,n} - Y_{k,n}^{(0)}||_2$. As $c_n \leqslant \sup_{k,n}(||y_{k,n}||_2 + \sup ||Y_{k,n}^{(0)}||_2) < \infty$ and we have the uniformly and geometrically NED property with NED coefficients $(\zeta^{1/\bar{d}_0})^s$.

(2) From Eq. (B.1), $||Y_1 - Y_2||_\infty \leqslant ||M_n||_\infty \cdot ||\eta_1 - \eta_2||_\infty$. By Prop. 1 in Jenish and Prucha (2012) and its proof, $||y_{i,n} - E[y_{i,n}|\mathcal{F}_{i,n}(s)]||_2 \leqslant \sigma_0 \sup_{i,n} \sum_{j:d(i,j)>s} m_{ij,n}$, thus it is sufficient to show that $\sup_{i,n} \sum_{j:d(i,j)>s} m_{ij,n} \leqslant C/s^{\alpha-d}$ for some constant $C > 0$.

For any positive integer l, define W_{An} as follows: when $w_{ij,n} \leqslant C_0(d(i,j)/l)^{-\alpha}$, where C_0 and α are those constants in Assumption 4, $w_{ij,An} = w_{ij,n}$; when $w_{ij,n} > C_0(d(i,j)/l)^{-\alpha}$, $w_{ij,An} = 0$. $W_{Bn} \equiv W_n - W_{An}$. Thus any element in W_{Bn} is either 0 or $> C_0(d(i,j)/l)^{-\alpha}$. Note that for all i's and j's, $w_{ij,An} w_{ij,Bn} = 0$. Now we calculate $\sum_{k_1} \cdots \sum_{k_{l-1}} w_{ik_1,n} w_{k_1k_2,n} \cdots w_{k_{l-1}j,n}$. For each product term of the summation, at least one element in a product is $\leqslant C_0/(d(i,j)/l)^\alpha$, because there exist at least two neighboring points in the chain $i \to k_1 \to \cdots k_{l-1} \to j$ such that their distance is at least $d(i,j)/l$. Thus $(W_{Bn}^l)_{ij} = \sum_{k_1} \cdots \sum_{k_{l-1}} w_{ik_1,Bn} w_{k_1k_2,Bn} \cdots w_{k_{l-1}j,Bn} = 0$. By Lemma A.4,

$$\begin{aligned}
&\sum_{k_1} \cdots \sum_{k_{l-1}} w_{ik_1,n} w_{k_1k_2,n} \cdots w_{k_{l-2}k_{l-1},n} w_{k_{l-1}j,n} \\
=&(W_n^l)_{ij} = [(W_{Bn} + W_{An})^l]_{ij} - (W_{Bn}^l)_{ij} \\
\leqslant& C_0(d(i,j)/l)^{-\alpha} \sum_{h=0}^{l-1} ||W_{Bn}||_\infty^h ||W_n^{l-h-1}||_1 \\
\leqslant& C_0(d(i,j)/l)^{-\alpha} \sum_{h=0}^{l-1} ||W_n||_\infty^h C_1(l-h-1)(\zeta/\lambda_m)^{l-h-1},
\end{aligned} \tag{B.3}$$

for some constant $C_1 > 0$, where, in the last inequality Lemma 1 is used for column sums. Thus, for any $j \neq i$,

$$(I_n - \lambda_m W_n)^{-1}_{ij} = \sum_{l=1}^{\infty}(\lambda_m W_n)^l_{ij} = \sum_{l=1}^{\infty}\lambda_m^l \sum_{k_1}\cdots\sum_{k_{l-1}} w_{ik_1,n}w_{k_1k_2,n}\cdots w_{k_{l-1}j,n}$$

$$\leqslant C_0C_1\sum_{l=1}^{\infty}\lambda_m^l(d(i,j)/l)^{-\alpha}\sum_{h=0}^{l-1}||W_n||_\infty^h(l-h-1)(\zeta/\lambda_m)^{l-h-1}$$

$$\leqslant C_0C_1\lambda_m\sum_{l=1}^{\infty}(d(i,j)/l)^{-\alpha}\sum_{h=0}^{l-1}(l-h-1)\zeta^{l-1}$$

$$=C_0C_1\lambda_m\sum_{l=1}^{\infty}(d(i,j)/l)^{-\alpha}\zeta^{l-1}l(l-1)/2$$

$$\leqslant C_0C_1\lambda_m d(i,j)^{-\alpha}\sum_{l=1}^{\infty}l^{2+\alpha}\zeta^{l-1}/2 = C_2d(i,j)^{-\alpha}$$

for some constant $C_2 > 0$. Recall $|\{j : m \leqslant d(i,j) < m+1\}| \leqslant C_3m^{d-1}$ for some constant $C_3 > 0$. Thus when s is large enough

$$\sup_{i,n}\sum_{j:d(i,j)>s} m_{ij,n} \leqslant \sup_{i,n}\sum_{m=[s]}^{\infty}\sum_{j:m\leqslant d(i,j)<m+1} C_3d(i,j)^{-\alpha}$$

$$\leqslant \sum_{m=[s]}^{\infty} C_3m^{d-1}C_2m^{-\alpha}$$

$$\leqslant \sum_{m=[s]}^{\infty} C_3C_2(m+1)^{d-1}[(m+1)/2]^{-\alpha}$$

$$\leqslant C_3C_22^{\alpha}\int_s^{\infty}x^{-\alpha+d-1}dx = C_3C_22^{\alpha}(\alpha-d)^{-1}s^{d-\alpha},$$

which implies that $||y_{i,n} - E[y_{i,n}|\mathcal{F}_{i,n}(s)]||_2 \leqslant \sigma_0C_3C_22^{\alpha}(\alpha-d)^{-1}/s^{\alpha-d}$. □

Proof of Corollary 1: (1) Assumptions 1 and 3 (1) imply $\sup_{i,n}|\{j : w_{i,j} \neq 0\}| < \infty$. So for $\sum_{j=1}^n w_{ij}y_{j,n}$, we are concerned only those j's with their locations within $\bar{d}_0$ from i because only such j's can satisfy $w_{ij,n} \neq 0$. Hence

$$||w_{i\cdot,n}Y_n - \mathrm{E}[w_{i\cdot,n}Y_n|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2 \leqslant \sum_{j=1}^n w_{ij,n}||y_{j,n} - \mathrm{E}[y_{j,n}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2$$

$$\leqslant \sum_{j=1}^n w_{ij,n}||y_{j,n} - \mathrm{E}[y_{j,n}|\mathcal{F}_{j,n}((m-1)\bar{d}_0)]||_2 \leqslant \sum_{j=1}^n w_{ij,n}C\zeta^{m-1} \leqslant C\lambda_m^{-1}\zeta^m.$$

(2) By Lemma A.1 in Jenish and Prucha (2009), $|\{j : m \leqslant d(i,j) < m+1\}| \leqslant C_1 m^{d-1}$ for some constant $C_1 > 0$. When s is large enough,

$$
\begin{aligned}
&||w_{i\cdot,n}Y_n - \mathrm{E}[w_{i\cdot,n}Y_n|\mathcal{F}_{i,n}(s)]||_2 \\
\leqslant & \sum_{k:d(k,i)\leqslant s/2} w_{ik,n}||y_{k,n} - \mathrm{E}[y_{k,n}|\mathcal{F}_{i,n}(s)]||_2 + \sum_{k:d(k,i)>s/2} w_{ik,n}||y_{k,n} - \mathrm{E}[y_{k,n}|\mathcal{F}_{i,n}(s)]||_2 \\
\leqslant & \sum_{k:d(k,i)\leqslant s/2} w_{ik,n}||y_{k,n} - \mathrm{E}[y_{k,n}|\mathcal{F}_{k,n}(s/2)]||_2 + \sum_{m=[s/2]}^{\infty} \sum_{k:m\leqslant d(k,i)<m+1} w_{ik,n}||y_{k,n}||_2 \\
\leqslant & \sum_{k:d(k,i)\leqslant s/2} w_{ik,n}C(s/2)^{d-\alpha} + (\sup_{n,k}||y_{k,n}||_2) \sum_{m=[s/2]}^{\infty} C_1 m^{d-1} C_0 m^{-\alpha} \\
\leqslant & ||W_n||_\infty C(s/2)^{d-\alpha} + (\sup_{n,k}||y_{k,n}||_2) C_1 C_0 2^{\alpha-d+1} \int_{[s/2]}^{\infty} \frac{dx}{x^{\alpha-d+1}} \leqslant C_2 s^{d-\alpha}
\end{aligned}
$$

for some constant $C_2 > 0$, where the second inequality comes from the fact that $\mathcal{F}_{k,n}(s/2) \subseteq \mathcal{F}_{i,n}(s)$ when $d(k,i) \leqslant s/2$. □

Proof of Lemma 3: Notice that $|z_1^2 - z_2^2| = |z_1 + z_2| \cdot |z_1 - z_2| \leqslant (|z_1| + |z_2| + 1) \cdot |z_1 - z_2|$, thus the conclusion holds by Lemmas 2 and A.5 with a there to be 1. □

Proof of Lemma 4: Let $G(x) \equiv \ln \Phi(x)$ and $g(x) = G'(x) = \frac{\phi(x)}{\Phi(x)} > 0$. Because $g'(x) = -\phi(x)[\phi(x) + x\Phi(x)]/\Phi^2(x)$ and it is known that $\phi(x) + x\Phi(x) > 0$, thus $g(x)$ is a strictly decreasing function. Notice that $\lim_{x\to-\infty} g(x) = \lim_{x\to-\infty} \frac{-x\phi(x)}{\phi(x)} = +\infty$. And $\lim_{x\to-\infty} g(x)/x = \lim_{x\to-\infty} \frac{\phi(x)}{\Phi(x)x} = \lim_{x\to-\infty} \frac{-x\phi(x)}{\Phi(x)+x\phi(x)} = \lim_{x\to-\infty} \frac{-1}{1+\Phi(x)/[x\phi(x)]} = -1$. Thus $|g(x)| < 2|x| + C_1$ for some constant $C_1 > 0$. By the mean value theorem for $G(x)$, there exists $\bar{x}$ between x and 0 such that $|G(x)| \leqslant |g(\bar{x})x| + |G(0)| \leqslant (2|x| + C_1)|x| + |G(0)| = 2x^2 + C_1|x| + |G(0)|$. Thus by Corollary 2, we have the uniform $L_{p/2}$ boundedness. Further, $|G(x_1) - G(x_2)| = |g(\bar{x})(x_1 - x_2)| \leqslant (2|x_1| + 2|x_2| + C_1)|x_1 - x_2|$, because this $\bar{x}$ lies between x_1 and x_2. Therefore, we obtain the conclusion by Corollary 2 and Lemma A.5. □

Proof of Lemma 5: Denote $g(x) = \phi(x)/\Phi(x)$. From the proof of Lemma 4, $|g(x)| < 2|x| + C_1$, thus $\{g(z_{i,n}(\theta))\}_{i=1}^{\infty}$ is uniformly L_p bounded. $g'(x) = -xg(x) - g^2(x)$. Because $g(x) \leqslant 2|x| + C_1$, $|xg(x)| \leqslant 2x^2 + C_1|x|$ and $g^2(x) \leqslant (2|x| + C_1)^2$. Therefore, $|g'(x)| \leqslant C_2(x^2 + 1)$ for some constant $C_2 > 0$. Hence, $|g(x_1) - g(x_2)| \leqslant C_2(x_1^2 + x_2^2 + 1)|x_1 - x_2|$. Then the NED properties of $\{g(z_{i,n})\}_{i=1}^{n}$ come from Corollary 2 and Lemma A.5. □

Proof of Lemma 6: Let $h(x) = \phi(x)x/\Phi(x)$ and $g(x) = \phi(x)/\Phi(x)$. From the proof of

Lemma 5, $g(x) < 2|x| + C_1$ for some constant $C_1 > 0$. Thus $|xg(x)| \leqslant 2x^2 + C_1|x|$ and by Lemma 2, $\{\phi(z_{i,n}(\theta))z_{i,n}(\theta)/\Phi(z_{i,n}(\theta))\}_{i=1}^n$ is uniformly $L_{p/2}$ bounded. Because $h'(x) = g(x) + xg'(x)$, $|h'(x)| \leqslant C_2(|x|^3+1)$ for some constant $C_2 > 0$. Then $|h(x_1)-h(x_2)| \leqslant C_2(|x_1|^3+|x_2|^3+1)|x_1-x_2|$. Hence, with Corollary 2 and Lemma A.5, we have the conclusion. □

Proof of Lemma 7: For any $\epsilon > 0$, let $B = \{|y_{i,n}^*| < \epsilon,\ |\mathrm{E}[y_{i,n}^*|\mathcal{F}_{i,n}(s)]| < \epsilon\}$. Since $|\mathbb{I}(x_1 > 0) - \mathbb{I}(x_2 > 0)| \leqslant \frac{|x_1-x_2|}{\epsilon}\mathbb{I}(|x_1| > \epsilon\, or\, |x_2| > \epsilon) + \mathbb{I}(|x_1| < \epsilon,\ |x_2| < \epsilon)$ (see the proof of Proposition 1 of Lei (2013)),

$$
\begin{aligned}
&||\mathbb{I}(y_{i,n}^* > 0) - \mathrm{E}[\mathbb{I}(y_{i,n}^* > 0)|\mathcal{F}_{i,n}(s)]||_2 \\
\leqslant& ||\mathbb{I}(y_{i,n}^* > 0) - \mathbb{I}\{\mathrm{E}[y_{i,n}^*|\mathcal{F}_{i,n}(s)] > 0\}||_2 \\
\leqslant& \left\{\int_{B^c}[y_{in}^* - \mathrm{E}(y_{in}^*|\mathcal{F}_{i,n}(s)]^2 dP/\epsilon^2 + P(B)\right\}^{1/2} \\
\leqslant& \frac{1}{\epsilon}||y_{i,n}^* - \mathrm{E}[y_{i,n}^*|\mathcal{F}_{i,n}(s)]||_2 + P(|y_{i,n}^*| < \epsilon)^{1/2} \\
\leqslant& \frac{1}{\epsilon}||y_{i,n}^* - \mathrm{E}[y_{i,n}^*|\mathcal{F}_{i,n}(s)]||_2 + (C_2\epsilon)^{1/2},
\end{aligned}
$$

for some constants $C_1 > 0$ and $C_2 > 0$. The last inequality comes from the uniform boundedness of the density function of $y_{i,n}^*$. Let $\epsilon = ||y_{i,n}^* - \mathrm{E}[y_{i,n}^*|\mathcal{F}_{i,n}(s)]||_2^{2/3}$, then we obtain the conclusion. □

Proof of Lemma 8: (i) Notice that

$$
(G_nA_nG_n)_{ii}^l = \sum_{j_1}\cdots\sum_{j_{l-1}} a_{ij_1,n}a_{j_1j_2,n}\cdots a_{j_{l-1}i,n}v_{i,n}^2v_{j_1,n}^2\cdots v_{j_{l-1},n}^2. \tag{B.4}
$$

When $a_{ij_1,n}a_{j_1j_2,n}\cdots a_{j_{l-1}i,n} \neq 0$, we have $d(i,j_1) \leqslant \bar{d}_0$, $d(j_1,j_2) \leqslant \bar{d}_0$, $\cdots$ Thus $\mathcal{F}_{j_h,n}((m-h)\bar{d}_0) \subseteq \mathcal{F}_{i,n}(m\bar{d}_0)$. For simplicity of notations, let $j_0 = i$. As the absolute values of $v_{j,n}$'s are less than or equal to one and the product of $v_{j,n}$'s is a Lipschitz function, when $m > l$, we have

$$
\begin{aligned}
&||v_{i,n}^2v_{j_1,n}^2\cdots v_{j_{l-1},n}^2 - \mathrm{E}[v_{i,n}^2v_{j_1,n}^2\cdots v_{j_{l-1},n}^2|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2 \\
\leqslant& \sum_{h=0}^{l-1}||v_{j_h,n}^2 - \mathrm{E}[v_{j_h,n}^2|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2 \\
\leqslant& 2\sum_{h=0}^{l-1}||v_{j_h,n} - \mathrm{E}[v_{j_h,n}|\mathcal{F}_{j_h,n}((m-h)\bar{d}_0)]||_2 \leqslant 2\sum_{h=0}^{l-1}C\eta^{m-h},
\end{aligned}
$$

where the second inequality follows from the fact that v^2 is also a Lipschitz function on $[-1,1]$.

When $m \leqslant h$, $|v_{j_h,n}^2 - \mathrm{E}[v_{j_h,n}^2|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2 \leqslant 1 \leqslant \max(1,C)\eta^{m-h}$. So the above inequality still holds if we replace C by $\max(1,C)$. Thus

$$\begin{aligned}
&||g_{i,ln} - \mathrm{E}[g_{i,ln}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2 \\
=&\sum_{j_1}\cdots\sum_{j_{l-1}} a_{ij_1,n}a_{j_1j_2,n}\cdots a_{j_{l-1}i,n}||v_{i,n}^2 v_{j_1,n}^2\cdots v_{j_{l-1},n}^2 - \mathrm{E}[v_{i,n}^2 v_{j_1,n}^2\cdots v_{j_{l-1},n}^2|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2 \\
\leqslant&||A_n||_\infty^l 2\max(1,C)\sum_{h=0}^{l-1}\eta^{m-h} \leqslant (2\max(1,C)\sum_{h=0}^{l-1}\eta^{l-h})\eta^m < (2\eta\max(1,C)/(1-\eta))\eta^m.
\end{aligned}$$

(ii) Notice

$$\begin{aligned}
&[(I_n - G_nA_nG_n)^{-1}G_nA_nG_n]_{ii}^l = \left(\prod_{j=1}^{l}\sum_{L_j=1}^{\infty}[G_nA_nG_n]^{L_j}\right)_{ii} \\
=&\sum_{k=l}^{\infty}\sum_{L_1+L_2+\cdots+L_l=k}[G_nA_nG_n]_{ii}^k = \sum_{k=l}^{\infty}\begin{pmatrix} k+l-1 \\ l-1 \end{pmatrix} g_{i,kn} \qquad \text{(B.5)} \\
\leqslant&\sum_{k=l}^{\infty}(k+l-1)^{l-1}g_{i,kn},
\end{aligned}$$

where the third equality follows from Sheldon (2002). When $m \leqslant l$,

$$\begin{aligned}
&||u_{i,n} - \mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2 \\
=&||\sum_{k=l}^{\infty}\sum_{L_1+L_2+\cdots+L_l=k}(G_nA_nG_n)_{ii}^k - \mathrm{E}[\sum_{k=l}^{\infty}\sum_{L_1+L_2+\cdots+L_l=k}(G_nA_nG_n)_{ii}^k|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2 \\
\leqslant&\sum_{k=l}^{\infty}(k+l-1)^{l-1}||g_{i,kn} - \mathrm{E}[g_{i,kn}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2 \\
\leqslant&\sum_{k=l}^{\infty}(k+l-1)^{l-1}\sum_{j_1}\cdots\sum_{j_{k-1}} a_{ij_1,n}\cdots a_{j_{k-1}i,n}||v_{i,n}^2\cdots v_{j_{k-1},n}^2 - \mathrm{E}[v_{i,n}^2\cdots v_{j_{k-1},n}^2|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2 \\
\leqslant&\sum_{k=l}^{\infty}(k+l-1)^{l-1}\eta^k < \int_{x=l-1}^{\infty}(x+l)^{l-1}\eta^x dx \leqslant \int_{x=m-1}^{\infty}(x+l)^{l-1}\eta^x dx,
\end{aligned}$$

where the third inequality comes from $||v_{i,n}^2\cdots v_{j_{k-1},n}^2 - \mathrm{E}[v_{i,n}^2\cdots v_{j_{k-1},n}^2|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2 \leqslant 1$. By

L'Hôpital's rule,

$$\begin{aligned}
&\lim_{m\to\infty}\frac{\int_{x=m-1}^{\infty}(x+l)^{l-1}\eta^x dx}{(m-1+l)^{l-1}\eta^{m-1}}\\
=&\lim_{m\to\infty}\frac{-(m-1+l)^{l-1}\eta^{m-1}}{(\ln\eta)(m-1+l)^{l-1}\eta^{m-1}+(l-1)(m-1+l)^{l-2}\eta^{m-1}}=\frac{-1}{\ln\eta}.
\end{aligned}$$

Thus, for some constant $\overline{C_{1l}}>0$, $\int_{x=m-1}^{\infty}(x+l)^{l-1}\eta^x dx\leqslant\overline{C_{1l}}m^l\eta^m$. Hence, $||u_{i,n}-\mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2<\overline{C_{1l}}m^l\eta^m$.

When $m>l$,

$$\begin{aligned}
&||u_{i,n}-\mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2\\
=&||\sum_{k=l}^{\infty}\sum_{L_1+L_2+\cdots+L_l=k}(G_nA_nG_n)^k_{ii}-\mathrm{E}[\sum_{k=l}^{\infty}\sum_{L_1+L_2+\cdots+L_l=k}(G_nA_nG_n)^k_{ii}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2\\
\leqslant&\sum_{k=l}^{m-1}(k+l-1)^{l-1}||g_{i,kn}-\mathrm{E}[g_{i,kn}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2\\
&+\sum_{k=m}^{\infty}(k+l-1)^{l-1}||g_{i,kn}-\mathrm{E}[g_{i,kn}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2\\
\leqslant&\sum_{k=l}^{m-1}(k+l-1)^{l-1}(2\eta C/(1-\eta))\eta^m+\sum_{k=m}^{\infty}(k+l-1)^{l-1}\eta^k\\
\leqslant&(2\eta C/(1-\eta))\int_{2l-1}^{m+l-2}x^{l-1}dx\cdot\eta^m+\sum_{k=m}^{\infty}(k+l-1)^{l-1}\eta^k\\
<&(2\eta C/(1-\eta))\cdot\frac{(m+l-2)^l-(2l-1)^l}{l}\eta^m+\int_{x=m-1}^{\infty}(x+l)^{l-1}\eta^x dx\\
\leqslant&\overline{C_{2l}}m^l\eta^m+\overline{C_{1l}}m^l\eta^m
\end{aligned}$$

for some constant $\overline{C_{2l}}>0$. Thus $||u_{i,n}-\mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(m\bar{d}_0)]||_2\leqslant(\overline{C_{1l}}+\overline{C_{2l}})m^l\eta^m$ holds for positive integers m. □

Proof of Lemma 9: (i) Given some distance $s>0$, we separate the product terms in the summation $\sum_{j_1}\cdots\sum_{j_{l-1}}$ into two parts: the first part, denoted as $P(1)$, with the distance of each pair of successive nodes in the chain $i\to j_1\to j_2\to\cdots\to j_{l-1}\to i$ is less than s/l , while the second part, denoted $P(2)$, consists of the other product terms. Thus in $P(2)$, there exists at least one element among $\{a_{ij_1,n},\ a_{j_1j_2,n},\cdots\ a_{j_{l-1}i,n}\}$ that is $\leqslant C_0(s/l)^{-\alpha}$. Let $j_0=i$. In $P(1)$, notice that

$$
\begin{aligned}
&||v_{i,n}^2 v_{j_1,n}^2 \cdots v_{j_{l-1},n}^2 - \mathrm{E}[v_{i,n}^2 v_{j_1,n}^2 \cdots v_{j_{l-1},n}^2|\mathcal{F}_{i,n}(s)]||_2 \\
\leqslant & \sum_{h=0}^{l-1} ||v_{j_h,n}^2 - \mathrm{E}[v_{j_h,n}^2|\mathcal{F}_{i,n}(s)]||_2 \leqslant \sum_{h=0}^{l-1} ||v_{j_h,n}^2 - \mathrm{E}^2[v_{j_h,n}|\mathcal{F}_{i,n}(s)]||_2 \\
\leqslant & 2\sum_{h=0}^{l-1} ||v_{j_h,n} - \mathrm{E}[v_{j_h,n}|\mathcal{F}_{j_h,n}(s-hs/l)]||_2 \leqslant 2C_1 l^p \sum_{h=0}^{l-1}(l-h)^{-p} \cdot s^{-p},
\end{aligned}
$$

where the third inequality follows from the fact that x^2 is a Lipschitz function on $[-1,1]$ and the fact that $B(j_h, s-hs/l) \subseteq B(i,s)$. Define A_{1n} as follows: when $a_{ij,n} \leqslant C_0(s/l)^{-\alpha}$, $a_{ij,1n} = a_{ij,n}$; when $a_{ij,n} > C_0(s/l)^{-\alpha}$, $a_{ij,1n} = 0$. $A_{2n} \equiv A_n - A_{1n}$. Thus every element in A_{2n} is either 0 or $> C_0(s/l)^{-\alpha}$. Hence,

$$
\begin{aligned}
&\sum_{P(2)} a_{ij_1,n}a_{j_1j_2,n}\cdots a_{j_{l-1}i,n} \leqslant [(A_{1n}+A_{2n})^l]_{ii} - (A_{2n}^l)_{ii} \\
\leqslant & C_0(s/l)^{-\alpha}\sum_{h=0}^{l-1}||A_{2n}||_\infty^h ||A_n^{l-h-1}||_1 \\
\leqslant & [C_0 l^\alpha \sum_{h=0}^{l-1}||A_n||_\infty^h C_2(l-h-1)\eta^{l-h-1}]s^{-\alpha} \\
\leqslant & [C_0C_2\eta^{l-1}l^\alpha\sum_{h=0}^{l-1}(l-h-1)]s^{-\alpha} \\
\leqslant & [C_0C_2\eta^{l-1}l^{\alpha+1}(l-1)/2]s^{-\alpha} = C_{2l}s^{-\alpha},
\end{aligned} \tag{B.6}
$$

where the first inequality follows from Lemma A.4. Hence,

$$
\begin{aligned}
&||g_{i,ln} - \mathrm{E}[g_{i,ln}|\mathcal{F}_{i,n}(s)]||_2 \\
\leqslant & \sum_{P(2)} a_{ij_1,n}a_{j_1j_2,n}\cdots a_{j_{l-1}i,n} + \sum_{P(1)} a_{ij_1,n}a_{j_1j_2,n}\cdots a_{j_{l-1}i,n}||v_{i,n}^2 v_{j_1,n}^2\cdots v_{j_{l-1},n}^2 \\
& - \mathrm{E}[v_{i,n}^2 v_{j_1,n}^2\cdots v_{j_{l-1},n}^2|\mathcal{F}_{i,n}(s)]||_2 \\
\leqslant & C_{2l}s^{-\alpha} + \eta^l \cdot 2C_1 l^p \sum_{h=0}^{l-1}(l-h)^{-p}\cdot s^{-p} \leqslant C_l s^{-p},
\end{aligned} \tag{B.7}
$$

where the last inequality results from $\alpha > p$.

(ii) By Eqs. (B.5), (B.6) and (B.7),

$$
\begin{aligned}
&||u_{i,n} - \mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(s)]||_2 \\
\leqslant &\sum_{k=l}^{\infty}(k+l-1)^{l-1}\left(2C_1k^p\sum_{h=0}^{k-1}(k-h)^{-p}\cdot\eta^k\cdot s^{-p} + [C_0C_2\eta^{k-1}k^{\alpha+1}(k-1)/2]s^{-\alpha}\right) \\
\leqslant &\sum_{k=l}^{\infty}(k+l-1)^{l-1}\left(2C_1k^{p+1}\cdot\eta^k\cdot s^{-p} + 0.5C_0C_2\eta^{k-1}k^{\alpha+2}\cdot s^{-\alpha}\right) \\
\leqslant &C_{3l}s^{-p} + C_{4l}s^{-\alpha} \leqslant \overline{C_l}s^{-p}
\end{aligned}
$$

for some constant $\overline{C_l} > 0$, because $0 < \eta < 1$. □

B.3 The Proof for Section 4

Proof of Lemma 10: Let $A = \{i : y_{i,n} > 0\}$ be the set of indexes under which $y_{i,n} > 0$ and $\mathbb{I}(A)$ be the event A's indicator. As Y_n is a random vector, each of its realizations gives a pattern of zero and positive observations. Each such pattern gives an A. Thus A represents a regime, and thus $\mathbb{I}(A)$ can be interpreted as a regime indicator. For each A, we may separate Y_n into two subvectors $Y_{1,n}$, whose elements are all zeros, and $Y_{2,n}$, whose elements are all positive. Similarly, $Y_n^{*'} = (Y_{1,n}^{*'}, Y_{2,n}^{*'})$ and

$$
W_n = \begin{pmatrix} W_{11,An} & W_{12,An} \\ W_{21,An} & W_{22,An} \end{pmatrix},
$$

so that $Y_{1,n}^* = \lambda_0 W_{12,An}Y_{2,n} + X_1\beta_0 + \epsilon_{1,n}$ and $Y_{2,n}^* = Y_{2,n} = \lambda_0 W_{22,An}Y_{2,n} + X_2\beta_0 + \epsilon_{2,n}$.

Next, we will calculate the marginal density function $f(y_{i,n}^*)$. As the range of $y_{i,n}^*$ is $(-\infty, +\infty)$, we discuss it in two cases: $y_{i,n}^*$ is positive or negative. In the following, "$-i$" means the rest $(n-1)$ elements without i.

When $y_{i,n}^* > 0$, there are 2^{n-1} possible different A's with $i \in A \subset \{1, 2, \cdots, n\}$. Given each A, $Y_{2,n}^* = \lambda_0 W_{22,An}Y_{2,n}^* + X_2\beta_0 + \epsilon_{2,n}$. Hence $Y_{2,n}^* = (I_{n_2} - \lambda_0 W_{22,An})^{-1}(X_2\beta_0 + \epsilon_{2,n})$. That is to say, on such a regime A, $y_{i,n}^* \sim N((I_{n_2} - \lambda_0 W_{22,An})_{i\cdot}^{-1}X_2\beta_0, \sigma_0((I_{n_2} - \lambda_0 W_{22,An})^{-1}(I_{n_2} - \lambda_0 W_{22,An}')^{-1})_{ii})$ and denote the corresponding density function as $f_A(y_{i,n}^*)$.

$$
f(y_{i,n}^*) = \mathbb{I}(y_{i,n}^* > 0)\sum_{i\in A\subset\{1,2,\cdots,n\}} f_A(y_{i,n}^*)\int f_A(Y_{-i,n}^*|y_{i,n}^*)dY_{-i,n}^*.
$$

Because the integral of a conditional density function is 1,

$$\sum_{i\in A\subset\{1,2,\cdots,n\}}\int f_A(Y^*_{-i,n}|y^*_{i,n})dY^*_{-i,n}=1.$$

Therefore, so long as we can show that $f_A(y^*_{i,n})$ is uniformly bounded, then $f(y^*_{i,n})$ is uniformly bounded on $y^*_{i,n}>0$. It suffices to show that $\inf_{A,i,n}((I_{n_2}-\lambda_0 W_{22,An})^{-1}(I_{n_2}-\lambda_0 W'_{22,An})^{-1})_{ii}>0$.

From Exercise 12.39 in Abadir and Magnus (2005), for any symmetric matrix M and compatible vector x, $x'Mx \geqslant \min eig(M)x'x$, where $\min eig(M)$ is the minimum characteristic root of M. Let $x=(0,\cdots,0,1,0,\cdots 0)$ in the above inequality, where 1 locates in the j^{th} position. We obtain $M_{jj}\geqslant \min eig(M)$, $\forall\, j$. Hence,

$$\begin{aligned}
&\inf_{A,i,n}((I_{n_2}-\lambda_0 W_{22,An})^{-1}(I_{n_2}-\lambda_0 W'_{22,An})^{-1})_{ii}\\
\geqslant\ &\inf_{A,i,n}\min eig((I_{n_2}-\lambda_0 W_{22,An})^{-1}(I_{n_2}-\lambda_0 W'_{22,An})^{-1})\\
=\ &\inf_{A,i,n}\min eig([(I_{n_2}-\lambda_0 W'_{22,An})(I_{n_2}-\lambda_0 W_{22,An})]^{-1})\\
=\ &\inf_{A,i,n}[\max eig((I_{n_2}-\lambda_0 W'_{22,An})(I_{n_2}-\lambda_0 W_{22,An}))]^{-1}\\
\geqslant\ &\inf_{A,i,n}[||(I_{n_2}-\lambda_0 W'_{22,An})(I_{n_2}-\lambda_0 W_{22,An})||_\infty]^{-1}\\
\geqslant\ &\inf_{A,i,n}[||I_{n_2}-\lambda_0 W'_{22,An}||_\infty\cdot||I_{n_2}-\lambda_0 W_{22,An}||_\infty]^{-1}\\
\geqslant\ &\inf_{A,i,n}[||I_{n_2}-\lambda_0 W_{22,An}||_1(1+\zeta)]^{-1}\geqslant[(1+\lambda_0\sup_n||W_n||_1)\cdot(1+\zeta)]^{-1}>0,
\end{aligned}$$

where the second inequality comes from the fact that the spectral radius of a matrix is less than or equal to any norm of that matrix and the last inequality is because $\sup_n||W_n||_1<\infty$ under either condition in Assumption 3.

When $y^*_{i,n}<0$, there are 2^{n-1} possible different A's where $A\subset\{1,2,\cdots,n\}\backslash\{i\}$. When $A=\emptyset$, $y_{j,n}=0$ for all j's, $Y^*_n=X_n\beta_0+\epsilon_n$. Thus, given $A=\emptyset$, the relevant density for $y^*_{i,n}$ takes the same form as the density of $N(0,\sigma_0^2)$. When $A \neq \emptyset$, because $Y_{2,n}^* = (I_{n_2} - \lambda W_{22,An})^{-1}(X_2\beta_0 + \epsilon_{2,n})$,

$$\begin{aligned}
&Y_{1,n}^* = \lambda_0 W_{12,An} Y_{2,n} + X_1\beta_0 + \epsilon_{1,n} \\
=&\lambda_0 W_{12,An}(I_{n_2} - \lambda_0 W_{22,An})^{-1}(X_2\beta_0 + \epsilon_{2,n}) + X_1\beta_0 + \epsilon_{1,n} \\
=&\lambda_0 W_{12,An}(I_{n_2} - \lambda_0 W_{22,An})^{-1} X_2\beta_0 + X_1\beta_0 + [\lambda_0 W_{12,An}(I_{n_2} - \lambda_0 W_{22,An})^{-1}\epsilon_{2,n} + \epsilon_{1,n}].
\end{aligned}$$

Thus, given A, the relevant density for y_{it}^* there takes the form as the density of $N(\lambda_0 w_{i\cdot,12,An}(I_{n_2} - \lambda_0 W_{22,An})^{-1}X_2\beta_0 + x_i\beta_0, \sigma_0^2 + \lambda_0^2\sigma_0^2 w_{i\cdot,12,An}(I_{n_2} - \lambda_0 W_{22,An})^{-1}[w_{i\cdot,12,An}(I_{n_2} - \lambda_0 W_{22,An})^{-1}]')$. Because

$$\sigma_0^2 + \lambda_0^2\sigma_0^2 w_{i\cdot,12,An}(I_{n_2} - \lambda_0 W_{22,An})^{-1}[w_{i\cdot,12,An}(I_{n_2} - \lambda_0 W_{22,An})^{-1}]' \geqslant \sigma_0^2,$$

$f(y_{i,n}^*)$ is also uniformly bounded when $y_{i,n}^* < 0$. □

Proof of Proposition 3: Let μ be a measure defined on $[0,\infty)$: $\mu([0,a]) = 1 + a$. And let $\mu^n = \mu \otimes \cdots \otimes \mu$ be the product measure of n $\mu's$. Because $\ln x \leqslant 2\sqrt{x} - 2$ for any $x \geqslant 0$, $\mathrm{E}\ln[L_n(\theta)/L_n(\theta_0)] \leqslant 2\mathrm{E}(\sqrt{L_n(\theta)/L_n(\theta_0)} - 1) = 2\int(\sqrt{L_n(\theta)/L_n(\theta_0)} - 1)L_n(\theta_0)d\mu^n(Y_n) = 2(\int \sqrt{L_n(\theta)L_n(\theta_0)}d\mu^n(Y_n) - 1) = -\int[\sqrt{L_n(\theta)} - \sqrt{L_n(\theta_0)}]^2 d\mu^n(Y_n) \leqslant 0$. Thus, $\mathrm{E}\ln L_n(\theta) \leqslant \mathrm{E}\ln L_n(\theta_0)$, and the equality holds if and only if $L_n(\theta) = L_n(\theta_0)$ μ^n-almost everywhere. If θ_0 is not identified, then there exists $\theta_1 \equiv (\lambda_1, \beta_1, \sigma_1) \neq \theta_0$ such that $\ln L_n(\theta_1) = \ln L_n(\theta_0)$ μ^n-almost everywhere, i.e.

$$\begin{aligned}
&\sum_{i=1}^n [1 - \mathbb{I}(y_{i,n} > 0)]\ln[1 - \Phi(\frac{\lambda_0}{\sigma_0} w_{i\cdot,n}Y_n + x_{i,n}\frac{\beta_0}{\sigma_0})] - \frac{1}{2}\ln(2\pi\sigma_0^2)\sum_{i=1}^n \mathbb{I}(y_{i,n} > 0) \\
&+ \ln|I_{2,n} - \lambda_0 W_{22,n}| - \frac{1}{2}\sum_{i=1}^n \mathbb{I}(y_{i,n} > 0)(\frac{1}{\sigma_0}y_{i,n} - \frac{\lambda_0}{\sigma_0}w_{i\cdot,n}Y_n - x_{i,n}\frac{\beta_0}{\sigma_0})^2 \\
=&\sum_{i=1}^n [1 - \mathbb{I}(y_{i,n} > 0)]\ln[1 - \Phi(\frac{\lambda_1}{\sigma_1} w_{i\cdot,n}Y_n + x_{i,n}\frac{\beta_1}{\sigma_1})] - \frac{1}{2}\ln(2\pi\sigma_1^2)\sum_{i=1}^n \mathbb{I}(y_{i,n} > 0) \\
&+ \ln|I_{2,n} - \lambda_1 W_{22,n}| - \frac{1}{2}\sum_{i=1}^n \mathbb{I}(y_{i,n} > 0)(\frac{1}{\sigma_1}y_{i,n} - \frac{\lambda_1}{\sigma_1}w_{i\cdot,n}Y_n - x_{i,n}\frac{\beta_1}{\sigma_1})^2
\end{aligned}$$

μ^n-almost everywhere. Because $P(y_{1,n} > 0, \cdots, y_{n,n} > 0) > 0$,

$$\begin{aligned} &\frac{n}{2}\ln(2\pi\sigma_0^2) - \ln|I_n - \lambda_0 W_n| + \frac{1}{2}\sum_{i=1}^n (\frac{1}{\sigma_0}y_{i,n} - \frac{\lambda_0}{\sigma_0}w_{i\cdot,n}Y_n - x_{i,n}\frac{\beta_0}{\sigma_0})^2 \\ =&\frac{n}{2}\ln(2\pi\sigma_1^2) - \ln|I_n - \lambda_1 W_n| + \frac{1}{2}\sum_{i=1}^n (\frac{1}{\sigma_1}y_{i,n} - \frac{\lambda_1}{\sigma_1}w_{i\cdot,n}Y_n - x_{i,n}\frac{\beta_1}{\sigma_1})^2 \end{aligned} \tag{B.8}$$

for $Y_n \in R^n_{++}$ almost everywhere. Differentiate the above equation with respect to $y_{j,n}$, we have

$$\begin{aligned} &-\frac{\lambda_0}{\sigma_0^2}\sum_{i=1}^n (y_{i,n} - \lambda_0 w_{i\cdot,n}Y_n - x_{i,n}\beta_0)w_{ij,n} + \frac{y_{j,n} - \lambda_0 w_{j\cdot,n}Y_n - x_{j,n}\beta_0}{\sigma_0^2} \\ =&-\frac{\lambda_1}{\sigma_1^2}\sum_{i=1}^n (y_{i,n} - \lambda_1 w_{i\cdot,n}Y_n - x_{i,n}\beta_1)w_{ij,n} + \frac{y_{j,n} - \lambda_1 w_{j\cdot,n}Y_n - x_{j,n}\beta_1}{\sigma_1^2}. \end{aligned} \tag{B.9}$$

Differentiate the above equation with respect to $y_{j,n}$ once more,

$$\frac{\lambda_0^2}{\sigma_0^2}\sum_{i=1}^n w_{ij,n}^2 + \frac{1 - 2\lambda_0 w_{jj,n}}{\sigma_0^2} = \frac{\lambda_1^2}{\sigma_1^2}\sum_{i=1}^n w_{ij,n}^2 + \frac{1 - 2\lambda_1 w_{jj,n}}{\sigma_1^2}.$$

Because $w_{jj,n} = 0$ and there exists $j \neq j'$ such that $\sum_{i=1}^n w_{ij,n}^2 \neq \sum_{i=1}^n w_{ij',n}^2$, we obtain $\lambda_0^2/\sigma_0^2 = \lambda_1^2/\sigma_1^2$ and $1/\sigma_0^2 = 1/\sigma_1^2$. Hence, $\sigma_0 = \sigma_1$ and $|\lambda_0| = |\lambda_1|$.

Differentiate Eq. (B.9) with respect to $y_{k,n}$ $(k \neq j)$,

$$\frac{\lambda_0^2}{\sigma_0^2}\sum_{i=1}^n w_{ik,n}w_{ij,n} - \frac{\lambda_0}{\sigma_0^2}(w_{kj,n} + w_{jk,n}) = \frac{\lambda_1^2}{\sigma_1^2}\sum_{i=1}^n w_{ik,n}w_{ij,n} - \frac{\lambda_1}{\sigma_1^2}(w_{kj,n} + w_{jk,n}).$$

Thus, $\lambda_0(w_{kj,n}+w_{jk,n}) = \lambda_1(w_{kj,n}+w_{jk,n})$. Because $W_n \neq 0$, and elements of W_n are non-negative, $\lambda_0 = \lambda_1$.

Eq. (B.9) implies that $\sum_{i=1}^n \lambda_0 w_{ij,n}x_{i,n}\beta_0 - x_j\beta_0 = \sum_{i=1}^n \lambda_0 w_{ij,n}x_{i,n}\beta_1 - x_j\beta_1$. Thus, $(I_n - \lambda_0 W_n')X_n\beta_0 = (I_n - \lambda_0 W_n')X_n\beta_1$. As $(I_n - \lambda_0 W_n')$ is invertible, $X_n\beta_0 = X_n\beta_1$. So, $\beta_0 = \beta_1$. □

Proof of Theorem 1: With Assumption 8, it is sufficient for us to show the uniform convergence in probability $\sup_{\theta\in\Theta}\frac{1}{n}|L_n(\theta) - \mathrm{E}L_n(\theta)| \xrightarrow{p} 0$ and the equicontinuity of $\{\mathrm{E}L_n(\theta)/n\}_{n=1}^\infty$.

The proof of $\sup_{\theta\in\Theta}\frac{1}{n}|L_n(\theta) - \mathrm{E}L_n(\theta)| \xrightarrow{p} 0$**:**

By Theorem 1 of Jenish and Prucha (2012), if a uniformly L_2-NED[11] random field is uniformly

[11]Theorem 1 of Jenish and Prucha (2012) is based on L_1-NED random fields, but since L_2-NED random fields are also L_1-NED random fields, their result is also applicable to L_2-NED random fields.

L_p bounded for some $p > 1$ and its base is independent,[12] then the weak law of large numbers (WLLN) holds. We have shown the uniform L_2-NED and uniform L_p boundedness properties of related terms of the Tobit model in the previous lemmas, propositions and corollaries.

$$\begin{aligned}
&\frac{1}{n}[\ln L_n(\theta) - \mathrm{E} \ln L_n(\theta)] \\
=&\frac{1}{n}\sum_{i=1}^{n}\{\mathbb{I}(y_{i,n}=0)\ln[1-\Phi(\frac{\lambda}{\sigma}w_{i\cdot,n}Y_n + x_{i,n}\frac{\beta}{\sigma})] - \mathrm{E}\{\mathbb{I}(y_{i,n}=0)\ln[1-\Phi(\frac{\lambda}{\sigma}w_{i\cdot,n}Y_n + x_{i,n}\frac{\beta}{\sigma})]\}\} \\
&-\frac{1}{2n}\ln(2\pi\sigma^2)\sum_{i=1}^{n}[\mathbb{I}(y_{i,n}>0) - \mathrm{E}\mathbb{I}(y_{i,n}>0)] + \frac{1}{n}(\ln|I_{2,n}-\lambda W_{22,n}| - \mathrm{E}\ln|I_{2,n}-\lambda W_{22,n}|)- \\
&\frac{1}{2n}\sum_{i=1}^{n}\{\mathbb{I}(y_{i,n}>0)(\frac{1}{\sigma}y_{i,n} - \frac{\lambda}{\sigma}w_{i\cdot,n}Y_n - x_{i,n}\frac{\beta}{\sigma})^2 - \mathrm{E}[\mathbb{I}(y_{i,n}>0)(\frac{1}{\sigma}y_{i,n} - \frac{\lambda}{\sigma}w_{i\cdot,n}Y_n - x_{i,n}\frac{\beta}{\sigma})^2]\}.
\end{aligned} \tag{B.10}$$

Because of the compactness of σ^2, the convergence to zero in probability of the second term on the right hand side of Eq. (B.10) will be uniform.

As $\{y_{i,n}\}$, $\{w_{i\cdot,n}Y_n\}$, $\{y_{i,n}^2\}$, $\{(w_{i\cdot,n}Y_n)^2\}$ and $\{\mathbb{I}(y_{i,n} > 0)\}$ are all uniformly L_2-NED and uniformly L_p bounded random fields for any natural number p, by Lemma A.2, their products remain uniformly L_2-NED and uniformly L_p bounded random fields. Thus, the pointwise WLLN is applicable to

$$\begin{aligned}
&\mathbb{I}(y_{i,n}>0)(\frac{1}{\sigma}y_{i,n} - \frac{\lambda}{\sigma}w_{i\cdot,n}Y_n - x_{i,n}\frac{\beta}{\sigma})^2 \\
=&\frac{1}{\sigma^2}\mathbb{I}(y_{i,n}>0)y_{i,n}^2 + \frac{\lambda^2}{\sigma^2}(w_{i\cdot,n}Y_n)^2\mathbb{I}(y_{i,n}>0) + \frac{1}{\sigma^2}(x_{i,n}\beta)^2\mathbb{I}(y_{i,n}>0) - \frac{2\lambda}{\sigma^2}w_{i\cdot,n}Y_n\mathbb{I}(y_{i,n}>0) \\
&-\frac{2\beta}{\sigma^2}x_{i,n}y_{i,n}\mathbb{I}(y_{i,n}>0) + \frac{2\lambda}{\sigma^2}(w_{i\cdot,n}Y_n)(x_{i,n}\beta)\mathbb{I}(y_{i,n}>0).
\end{aligned}$$

Further, because of compactness of the parameter space, $1/\sigma^2$, λ^2/σ^2, β^2/σ^2, λ/σ^2, β/σ^2 and $\beta\lambda/\sigma^2$ are all bounded, uniform convergence in probability follows.

Now we will show the uniform convergence of the first term, $L_{1n}(\theta) \equiv \frac{1}{n}\sum_{i=1}^{n}\{\mathbb{I}(y_{i,n}=0)\ln[1-\Phi(\frac{\lambda}{\sigma}w_{i\cdot,n}Y_n + x_{i,n}\frac{\beta}{\sigma})] - \mathrm{E}\{\mathbb{I}(y_{i,n}=0)\ln[1-\Phi(\frac{\lambda}{\sigma}w_{i\cdot,n}Y_n + x_{i,n}\frac{\beta}{\sigma})]\}\}$, in Eq. (B.10). For any $\theta \in \Theta$, by Lemmas 4 and 7, $L_{1n} \xrightarrow{p} 0$. By the compactness of the parameter space and Theorem 1 in Andrews (1992), it is sufficient to show that $L_{1n}(\theta)$ is stochastically equicontinuous. To do so, we only need

[12] More generally, the base is a spatial mixing process in Jenish and Prucha (2002). For our purpose, we focus on an independent base, which is a special case.

to check the conditions of Corollary 3.1 in Andrews (1992). Let $\tilde{\lambda} = \lambda/\sigma$, $\tilde{\beta} = \beta/\sigma$, $\tilde{\sigma} = \sigma^{-1}$, which is similar to the reparameterization due to Olsen (1978), and then

$$\frac{\partial \tilde{\theta}}{\partial \theta'} = \begin{pmatrix} \sigma^{-1} & 0 & -\lambda/\sigma^2 \\ 0 & \sigma^{-1} I_K & -\beta/\sigma^2 \\ 0 & 0 & -\sigma^{-2} \end{pmatrix}.$$

Because the parameter space is compact, $|\partial \tilde{\theta}_j / \partial \theta_k|$ is bounded for all j's and k's. Then $L_{1n}(\tilde{\lambda}, \tilde{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\{\mathbb{I}(y_{i,n} = 0)\ln[1 - \Phi(\tilde{\lambda} w_{i\cdot,n} Y_n + x_{i,n}\tilde{\beta})] - \mathrm{E}\{\mathbb{I}(y_{i,n} = 0)\ln[1 - \Phi(\tilde{\lambda} w_{i\cdot,n} Y_n + x_{i,n}\tilde{\beta})]\}\}$. By Lemma A.6, it is sufficient to show that $L_{1n}(\tilde{\lambda}, \tilde{\beta})$ is stochastically equicontinuous. Evidently, the ranges of $\tilde{\lambda}$ and $\tilde{\beta}$ are compact in their parameter spaces. Denote $\tilde{\lambda}_m = \sup \tilde{\lambda}$ and $\tilde{\beta}_m = \sup_{\tilde{\beta}} \max_{j=1}^{K} |\tilde{\beta}_j|$ from their compact parameter spaces. Recall from the proof of Lemma 4, $|\ln \Phi(x_1) - \ln \Phi(x_2)| \leqslant (2|x_1| + 2|x_2| + C_1)|x_1 - x_2|$ for some constant C_1. Hence,

$$\begin{aligned}
&|\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(y_{i,n} = 0)\{\ln \Phi(-\tilde{\lambda} w_{i\cdot,n} Y_n - x_{i,n}\tilde{\beta}) - \ln \Phi(-\tilde{\lambda}' w_{i\cdot,n} Y_n - x_{i,n}\tilde{\beta}')\}| \\
\leqslant & \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(y_{i,n} = 0)[2|\tilde{\lambda} w_{i\cdot,n} Y_n + x_{i,n}\tilde{\beta}| + 2|\tilde{\lambda}' w_{i\cdot,n} Y_n + x_{i,n}\tilde{\beta}'| + C_1]|(\tilde{\lambda} - \tilde{\lambda}')w_{i\cdot,n} Y_n + x_{i,n}(\tilde{\beta} - \tilde{\beta}')| \\
\leqslant & \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(y_{i,n} = 0)[4\tilde{\lambda}_m |w_{i\cdot,n} Y_n| + 4\tilde{\beta}_m \sum_{j=1}^{K} |x_{ij,n}| + C_1](|w_{i\cdot,n} Y_n| + \sum_{j=1}^{K} |x_{ij,n}|) \cdot (|\tilde{\lambda} - \tilde{\lambda}'| + |\tilde{\beta} - \tilde{\beta}'|).
\end{aligned}$$

By Assumption 3 and Lemma 2, $\sup_{i,n} ||w_{i\cdot,n} Y_n|||_p < \infty$ for any natural number p. Hence $\sup_{i,n} ||4\tilde{\lambda}_m |w_{i\cdot,n} Y_n| + 4\tilde{\beta}_m \sum_{j=1}^{K} |x_{ij,n}| + C_1||_p < \infty$ and $\sup_{i,n} ||(|w_{i\cdot,n} Y_n| + \sum_{j=1}^{K} |x_{ij,n}|)||_p < \infty$. Thus, we have $\sup_{i,n} ||\{4\tilde{\lambda}_m |w_{i\cdot,n} Y_n| + 4\tilde{\beta}_m \sum_{j=1}^{K} |x_{ij,n}| + C_1\}(|w_{i\cdot,n} Y_n| + \sum_{j=1}^{K} |x_{ij,n}|)||_p < \infty$ by Cauchy's inequality. Further, by Minkowski's inequality, $\sup_n ||\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}(y_{i,n} = 0)[4\tilde{\lambda}_m |w_{i\cdot,n} Y_n| + 4\tilde{\beta}_m \sum_{j=1}^{K} |x_{ij,n}| + C_1](|w_{i\cdot,n} Y_n| + \sum_{j=1}^{K} |x_{ij,n}|)||_p < \infty$. Therefore, by Lemma 1 (a) in Andrews (1992), $L_{1n}(\theta)$ is stochastically equicontinuous and $\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\mathbb{I}(y_{i,n} = 0)\ln \Phi(-\tilde{\lambda} w_{i\cdot,n} Y_n - x_{i,n}\tilde{\beta})]\}$ is equicontinuous.

Now, it remains to show the uniform convergence of $\frac{1}{n}(\ln |I_{2,n} - \lambda W_{22,n}| - \mathrm{E}\ln |I_{2,n} - \lambda W_{22,n}|)$. We will use a strategy in Qu and Lee (2013a). Let $G_n(Y_n) = diag(\mathbb{I}(y_{1,n} > 0), \cdots, \mathbb{I}(y_{n,n} > 0))$.

Then as shown in Qu and Lee (2013a), we have

$$
\begin{aligned}
&\ln|I_{2,n}-\lambda W_{22,n}| = -\sum_{l=1}^{\infty}\frac{\lambda^l}{l}tr\{[G_n(Y_n)W_nG_n(Y_n)]^l\} \\
=&-\sum_{i=1}^{n}\mathbb{I}(y_{i,n}>0)\left(\sum_{l=1}^{\infty}\frac{\lambda^l}{l}[G_n(Y_n)W_nG_n(Y_n)]^l\right)_{ii} \\
=&-\sum_{i=1}^{n}\mathbb{I}(y_{i,n}>0)\sum_{l=1}^{\infty}\frac{\lambda^{l+1}}{l+1}\sum_{j_1}\cdots\sum_{j_l}w_{ij_1,n}w_{j_1j_2,n}\cdots w_{j_{l-1}j_l,n}w_{j_li,n}\mathbb{I}(y_{j_1,n}>0)\cdots\mathbb{I}(y_{j_l,n}>0).
\end{aligned}
$$

For any $\epsilon > 0$, let K_ϵ be a natural number that does not depend on n and its value will be determined later. Divide the summation over l into two parts: $S_{i,n}(\lambda) \equiv \sum_{l=1}^{K_\epsilon}\frac{\lambda^{l+1}}{l+1}g_{il,n}$, where $g_{il,n} = \mathbb{I}(y_{i,n} > 0)\sum_{j_1}\cdots\sum_{j_l}w_{ij_1,n}w_{j_1j_2,n}\cdots w_{j_li,n}\mathbb{I}(y_{j_1,n} > 0)\cdots\mathbb{I}(y_{j_l,n} > 0)$, and $R_{i,n}(\lambda) \equiv \mathbb{I}(y_{i,n} > 0)\sum_{l=K_\epsilon+1}^{\infty}\frac{\lambda^{l+1}}{l+1}\sum_{j_1}\cdots\sum_{j_l}w_{ij_1,n}w_{j_1j_2,n}\cdots w_{j_li,n}\mathbb{I}(y_{j_1,n} > 0)\cdots\mathbb{I}(y_{j_l,n} > 0)$. We will show that $\sup_{\lambda\in\Lambda}|\frac{1}{n}\sum_{i=1}^{n}[S_{i,n}(\lambda)-\mathrm{E}S_{i,n}(\lambda)]| \xrightarrow{p} 0$ and $\sup_{\lambda,n}|\frac{1}{n}\sum_{i=1}^{n}[R_{i,n}(\lambda)-\mathrm{E}R_{i,n}(\lambda)]| < \epsilon/2$. From Lemma (9), for each natural number $l \leqslant K_\epsilon$, $\{g_{il,n}\}$ is a uniform NED random field. Furthermore, from the definition of $g_{il,n}$, we have $\sup_{i,n}|g_{il,n}| < \infty$. Then $\frac{1}{n}\sum_{i=1}^{n}(g_{il,n}-\mathrm{E}g_{il,n}) \xrightarrow{p} 0$ follows from the WLLN in Jenish and Prucha (2012). Thus we have shown that $\sup_{\lambda\in\Lambda}|\frac{1}{n}\sum_{i=1}^{n}[S_{i,n}(\lambda) - \mathrm{E}S_{i,n}(\lambda)]| \xrightarrow{p} 0$.

Notice that there is a constant K_ϵ such that

$$
\sup_{\lambda\in\Lambda}|\frac{1}{n}\sum_{i=1}^{n}R_{i,n}(\lambda)| \leqslant \sum_{l=K_\epsilon+1}^{\infty}\frac{\lambda_m^l}{l}||W_n||_\infty^{l+1} < \frac{\zeta^{K_\epsilon+2}}{\lambda_m(1-\zeta)} < \frac{\epsilon}{4}.
$$

Similarly, we also have $\sup_{\lambda\in\Lambda}|\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}R_{i,n}(\lambda)| < \frac{\epsilon}{4}$. Hence, $\sup_{\lambda\in\Lambda}|\frac{1}{n}\sum_{i=1}^{n}[R_{i,n}(\lambda)-\mathrm{E}R_{i,n}(\lambda)]| < \frac{\epsilon}{2}$. Therefore,

$$
\begin{aligned}
&P(\sup_{\lambda\in\Lambda}\frac{1}{n}\Big|\ln|I_{2,n}-\lambda W_{22,n}| - \mathrm{E}\ln|I_{2,n}-\lambda W_{22,n}|\Big| > \epsilon) \\
\leqslant& P(\sup_{\lambda\in\Lambda}|\frac{1}{n}\sum_{i=1}^{n}[S_{i,n}(\lambda)-\mathrm{E}S_{i,n}(\lambda)]| + \sup_{\lambda\in\Lambda}|\frac{1}{n}\sum_{i=1}^{n}[R_{i,n}(\lambda)-\mathrm{E}R_{i,n}(\lambda)]| > \epsilon) \\
\leqslant& P(\sup_{\lambda\in\Lambda}|\frac{1}{n}\sum_{i=1}^{n}[S_{i,n}(\lambda)-\mathrm{E}S_{i,n}(\lambda)]| > \epsilon/2) \to 0.
\end{aligned}
$$

The proof of the equicontinuity of $\{\mathrm{E}L_n(\theta)/n\}_{n=1}^{\infty}$:

In the proof of the uniform convergence in probability, we have shown that $\{\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\mathbb{I}(y_{i,n} =$

$0)\ln\Phi(-\tilde{\lambda}w_{i\cdot,n}Y_n - x_{i,n}\tilde{\beta})]\}$ is equicontinuous. Because of the compactness of the parameter space, the equicontinuity of $\frac{1}{2n}\ln(2\pi\sigma^2)\mathrm{E}\sum_{i=1}^n \mathbb{I}(y_{i,n}>0)$ is obvious. We still need to show the other two terms, $\frac{1}{n}\mathrm{E}\ln|I_{2,n}-\lambda W_{22,n}|$ and $\frac{1}{2n}\mathrm{E}\sum_{i=1}^n \mathbb{I}(y_{i,n}>0)(\frac{1}{\sigma}y_{i,n}-\frac{\lambda}{\sigma}w_{i\cdot,n}Y_n - x_{i,n}\frac{\beta}{\sigma})^2$, in $\mathrm{E}L_n(\theta)/n$ are equicontinuous. By Lemma A.6, we only need to show that $\frac{1}{2n}\mathrm{E}\sum_{i=1}^n \mathbb{I}(y_{i,n}>0)(\tilde{\sigma}y_{i,n}-\tilde{\lambda}w_{i\cdot,n}Y_n - x_{i,n}\tilde{\beta})^2$ is equicontinuous with respect to $\tilde{\theta}$.

$$
\begin{aligned}
&\frac{1}{2n}\mathrm{E}\sum_{i=1}^n \mathbb{I}(y_{i,n}>0)(\tilde{\sigma}y_{i,n}-\tilde{\lambda}w_{i\cdot,n}Y_n - x_{i,n}\tilde{\beta})^2\\
=&\frac{1}{2n}\mathrm{E}\sum_{i=1}^n \mathbb{I}(y_{i,n}>0)[\tilde{\sigma}^2y_{i,n}^2+\tilde{\lambda}^2(w_{i\cdot,n}Y_n)^2+\tilde{\beta}'x_{i,n}'x_{i,n}\tilde{\beta}\\
&-2\tilde{\sigma}\tilde{\lambda}y_{i,n}w_{i\cdot,n}Y_n - 2\tilde{\sigma}y_{i,n}x_{i,n}\tilde{\beta}+2\tilde{\lambda}w_{i\cdot,n}Y_nx_{i,n}\tilde{\beta}].
\end{aligned}
$$

Because $\{y_{i,n}^2\}_{i=1}^n$, $\{(w_{i\cdot,n}Y_n)^2\}_{i=1}^n$ and $\{y_{i,n}w_{i\cdot,n}Y_n\}_{i=1}^n$ are uniformly L_p bounded for any natural number p, $\{x_{i,n}\}_{i=1}^n$ is uniformly bounded, and the parameter space is compact, we have the equicontinuity of $\frac{1}{2n}\mathrm{E}\sum_{i=1}^n \mathbb{I}(y_{i,n}>0)(\tilde{\sigma}y_{i,n}-\tilde{\lambda}w_{i\cdot,n}Y_n - x_{i,n}\tilde{\beta})^2$.

Recall $\mathrm{E}\ln|I_{2,n}-\lambda W_{22,n}| = \mathrm{E}\ln|I_n-\lambda\widetilde{W_n}|$. Hence, the equicontinuity of $\mathrm{E}\ln|I_{2,n}-\lambda W_{22,n}|/n$ comes from

$$
\begin{aligned}
&\left|\frac{1}{n}[\mathrm{E}\ln|I_{2,n}-\lambda_1 W_{22,n}| - \mathrm{E}\ln|I_{2,n}-\lambda_2 W_{22,n}|]\right|\\
\leqslant&\frac{1}{n}|\mathrm{E}tr[(I_n-\bar{\lambda}\widetilde{W_n})^{-1}\widetilde{W_n}(\lambda_1-\lambda_2)]|\\
\leqslant&\sup_n ||(I_n-\bar{\lambda}\widetilde{W_n})^{-1}\widetilde{W_n}||_\infty\cdot|\lambda_1-\lambda_2|\\
\leqslant&\sup_n\sum_{l=0}^\infty \lambda_m^l||\widetilde{W_n}||_\infty^{l+1}\cdot|\lambda_1-\lambda_2|\\
\leqslant&\lambda_m^{-1}\sum_{l=0}^\infty\zeta^{l+1}\cdot|\lambda_1-\lambda_2| = \frac{\lambda_m^{-1}\zeta}{1-\zeta}|\lambda_1-\lambda_2|.
\end{aligned}
$$

□

B.4 The Proof of Section 5

Proof of Lemma 11: By Corollary 1 in Jenish and Prucha (2012), with Assumption 10, to show the central limit theorem, it is sufficient to check the uniform $L_{2+\delta}$ integrability, for some $\delta>0$, the uniform NED property of $\{||q_{i,n}(\theta_0)||\}_{i=1}^n$ in Eq. (5.1), where $||\cdot||$ is the Euclidean vector norm, and the decreasing rate of the NED coefficient.

We discuss the NED property separately under the two different settings in Assumption 3. Under Assumption (3)(1), $\{z_{i,n}(\theta)^2\}_{i=1}^n$, $\{\ln\Phi(z_{i,n}(\theta))\}_{i=1}^n$, $\{\phi(z_{i,n}(\theta))/\Phi(z_{i,n}(\theta))\}$, $\{\mathbb{I}(y_{i,n}=0)\}_{i=1}^n$, $\{\phi(z_{i,n}(\theta))w_{i\cdot,n}Y_n/\Phi(z_{i,n}(\theta))\}$ and $\{r_{i,n}(\lambda_0)\}_{i=1}^n$ are uniformly and geometrically L_2-NED random fields. Thus, by Lemma A.2, their products are also uniformly and geometrically L_2-NED random fields. That is to say, all terms in $q_{i,n}(\theta_0)$ are L_2-NED random fields. By Lemma B.4 in Xu and Lee (2013) for the Euclidean norm, $\{||q_{i,n}(\theta_0)||\}_{i=1}^n$ is a uniformly and geometrically L_2-NED random field. Then (c) and (d) in Assumption 4 in Jenish and Prucha (2012) are satisfied.

Under Assumption (3)(2), from Proposition 2 (2), $\{r_{i,n}(\lambda_0)\}_{i=1}^n$ is a uniformly L_2-NED random field with coefficient $1/s^{(\alpha-d)/3}$. Because all moments exist for a normal distribution, the p in Lemmas 3-6 can be arbitrarily large. As $\lim_{p\to\infty}(p-4)/(2p-4)=1/2$ and $\lim_{p\to\infty}(p-8)/(2p-8)=1/2$, $\{z_{i,n}(\theta)^2\}_{i=1}^n$, $\{\ln\Phi(z_{i,n}(\theta))\}_{i=1}^n$, $\{\phi(z_{i,n}(\theta))/\Phi(z_{i,n}(\theta))\}$ and $\{\phi(z_{i,n}(\theta))w_{i\cdot,n}Y_n/\Phi(z_{i,n}(\theta))\}$ are all uniformly L_2-NED random fields with NED coefficient $(1/s^{\alpha-d})^{(1-\gamma)/2}$ for any $0<\gamma<1$. From Lemma 7, $\{\mathbb{I}(y_{i,n}=0)\}_{i=1}^n$ is a uniformly NED random field with coefficient $1/s^{(\alpha-d)/3}$. Similarly, $\lim_{r\to\infty}(r-2)/(2r-2)=1/2$. Thus, by Lemma A.2, where the r can be arbitrarily large, $\{\mathbb{I}(y_{i,n}=0)\phi(z_{i,n})w_{i\cdot,n}Y_n/\Phi(z_{i,n})\}_{i=1}^n$ is a uniformly NED random field with NED coefficient $(1/s^{\alpha-d})^{(1-\gamma)/6}$ for any $0<\gamma<1$. Similarly, $\{\mathbb{I}(y_{i,n}=0)z_{i,n}w_{i\cdot,n}Y_n\}_{i=1}^n$, $\{\mathbb{I}(y_{i,n}>0)z_{i,n}x_{ik}\}_{i=1}^n$, $\{\mathbb{I}(y_{i,n}=0)\phi(z_{i,n})x_{ik,n}/\Phi(z_{i,n})\}_{i=1}^n$, $\{\mathbb{I}(y_{i,n}>0)x_{ik,n}z_{i,n}\}_{i=1}^n$, $\{\mathbb{I}(y_{i,n}>0)z_{i,n}^2\}_{i=1}^n$ are all uniformly L_2-NED random fields with NED coefficient $(1/s^{\alpha-d})^{(1-\gamma)/6}$, for any $0<\gamma<1$. Thus all the terms in the score are uniformly NED random fields and the slowest NED coefficient is $(1/s^{\alpha-d})^{(1-\gamma)/6}$. Hence, for any $0<\gamma<1$, $\{||q_{i,n}(\theta_0)||\}_{i=1}^n$ is a uniformly L_2-NED random field with coefficient $(1/s^{\alpha-d})^{(1-\gamma)/6}$. The rate condition $\sum_{s=1}^{\infty}(1/s^{\alpha-d})^{(1-\gamma)/6}s^{d-1}<\infty$ for some $0<\gamma<1$ is satisfied, when $\alpha>7d$.

Next, it remains to check the uniform $L_{2+\delta}$ integrability of the Euclidean norm of $\{q_{i,n}(\theta_0)\}_{i=1}^n$, denoted $\{||q_{i,n}(\theta_0)||\}_{i=1}^n$, for some $\delta>0$. It is sufficient to show that $\sup_{i,n}\mathrm{E}||q_{i,n}(\theta)||^{2+\delta}<\infty$ (see Exercise 5.4, page 54, Shorack, 2000). From the C_r-inequality[13],

$$\mathrm{E}||q_{i,n}(\theta)||^{2+\delta}\leqslant(2+K)^{\delta/2}\{\mathrm{E}|\frac{\partial\ln L_n(\theta)}{\partial\lambda}|^{2+\delta}+\sum_{k=1}^{K}\mathrm{E}|\frac{\partial\ln L_n(\theta)}{\partial\beta_k}|^{2+\delta}+\mathrm{E}|\frac{\partial\ln L_n(\theta)}{\partial\sigma}|^{2+\delta}\}.$$

[13]If $r>1$, then $\mathrm{E}|X_1+\cdots+X_k|^r\leqslant k^{r-1}(\mathrm{E}|X_1|^r+\cdots+\mathrm{E}|X_k|^r)$

It is enough to show the uniform $L_{2+\delta}$ integrability of each component on the right hand side of the above inequality, again by the C_r-inequality.

In previous lemmas, we have shown the uniform L_p boundedness of $\{y_{i,n}\}_{i=1}^n$, $\{w_{i\cdot,n}Y_n\}_{i=1}^n$ and $\{\phi(\frac{\lambda}{\sigma}w_{i\cdot,n}Y_n + x_{i,n}\frac{\beta}{\sigma})/\Phi(-\frac{\lambda}{\sigma}w_{i\cdot,n}Y_n - x_{i,n}\frac{\beta}{\sigma})\}_{i=1}^n$, for any natural number p. By Cauchy's inequality, we know that all of their pairwise products are also uniformly L_p bounded. Furthermore, $\{r_{i,n}(\lambda)\}_{i=1}^n$ is uniformly bounded because $|r_{i,n}(\lambda)| \leqslant \sum_{l=1}^{\infty}|\lambda|^l \sum_{j_1}\cdots\sum_{j_l} w_{ij_1,n}w_{j_1j_2,n}\cdots w_{j_li,n} \leqslant \sum_{l=1}^{\infty}|\lambda|^l||W_n||_{\infty}^{l+1} \leqslant \lambda_m^{-1}\zeta^2/(1-\zeta)$. Thus all the conditions for the CLT of Jenish and Prucha (2012) are satisfied, and the asymptotic normality of the normalized score vector follows. □

Proof of Theorem 2: Because $0 = \frac{\partial \ln L_n(\hat{\theta})}{\partial\theta} = \frac{\partial \ln L_n(\theta_0)}{\partial\theta} + \frac{\partial^2 \ln L_n(\bar{\theta})}{\partial\theta\partial\theta'}\cdot(\hat{\theta}-\theta_0)$, $\sqrt{n}(\hat{\theta}-\theta_0) = [\frac{1}{n}\frac{\partial^2 \ln L_n(\bar{\theta})}{\partial\theta\partial\theta'}]^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta}$.

Claim: For any consistent estimate $\bar{\theta}$ of θ_0, $\frac{1}{n}[\frac{\partial^2 \ln L_n(\bar{\theta})}{\partial\theta\partial\theta'} - \mathrm{E}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}] = o_p(1)$.

As $\mathrm{E}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} = -\mathrm{Var}\sum_{i=1}^n q_{i,n}(\theta_0)$, with the above claim, and the asymptotic normality of the normalized score vector, we have the asymptotic normality of the MLE: $\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,\Sigma_0^{-1})$.

In the following paragraphs, we will prove the above claim on the second order derivatives of the log likelihood function. Define $\psi(x)$ as

$$\psi(x) = \frac{d[\phi(x)/\Phi(x)]}{dx} = \frac{-x\phi(x)}{\Phi(x)} - \frac{\phi^2(x)}{\Phi^2(x)}.$$

Then the second derivatives are

$$\begin{aligned}\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda^2} =& \sum_{i=1}^n \mathbb{I}(y_{i,n}=0)\psi(z_{i,n}(\theta))\left(\frac{w_{i\cdot,n}Y_n}{\sigma}\right)^2 - tr[(I_{2,n}-\lambda W_{22,n})^{-1}W_{22,n}]^2\\ &-\sum_{i=1}^n \mathbb{I}(y_{i,n}>0)\left(\frac{w_{i\cdot,n}Y_n}{\sigma}\right)^2\end{aligned}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\beta} = \sum_{i=1}^n \mathbb{I}(y_{i,n}=0)\psi(z_{i,n}(\theta))\frac{w_{i\cdot,n}Y_n}{\sigma}\frac{x'_{i,n}}{\sigma} - \sum_{i=1}^n \mathbb{I}(y_{i,n}>0)\frac{w_{i\cdot,n}Y_n}{\sigma}\frac{x'_{i,n}}{\sigma}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\sigma} = \sum_{i=1}^n \mathbb{I}(y_{i,n}=0)[\frac{\phi(z_{i,n}(\theta))w_{i\cdot,n}Y_n}{\Phi(z_{i,n}(\theta))\sigma^2} - \psi(z_{i,n})\frac{\lambda w_{i\cdot,n}Y_n + x_{i,n}\beta}{\sigma^2}\frac{w_{i\cdot,n}Y_n}{\sigma}]$$
$$- 2\sum_{i=1}^n \mathbb{I}(y_{i,n}>0)\frac{y_{i,n} - \lambda w_{i\cdot,n}Y_n - x_{i.n}\beta}{\sigma^3}w_{i\cdot,n}Y_n,$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\beta'} = \sum_{i=1}^n \mathbb{I}(y_{i,n}=0)\psi(z_{i,n}(\theta))\frac{x'_{i,n}x_{i,n}}{\sigma^2} - \sum_{i=1}^n \mathbb{I}(y_{i,n}>0)\frac{x'_{i,n}x_{i,n}}{\sigma^2},$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\sigma} = \sum_{i=1}^n \mathbb{I}(y_{i,n}=0)\left[\frac{\phi(z_{i,n})}{\Phi(z_{i,n})} - \psi(z_{i,n})\frac{\lambda w_{i\cdot,n}Y_n + x_{i,n}\beta}{\sigma}\right]\frac{x'_{i,n}}{\sigma^2}$$
$$- 2\sum_{i=1}^n \mathbb{I}(y_{i,n}>0)\sigma^{-3}(y_{i,n} - \lambda w_{i\cdot,n}Y_n - x_{i.n}\beta)x'_{i,n},$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\sigma\partial\sigma} = \sum_{i=1}^n \mathbb{I}(y_{i,n}=0)[\psi(z_{i,n})\left(\frac{\lambda w_{i\cdot,n}Y_n + x_{i.n}\beta}{\sigma^2}\right)^2 - 2\frac{\phi(z_{i,n}(\theta))(\lambda w_{i\cdot,n}Y_n + x_{i.n}\beta)}{\Phi(z_{i,n}(\theta))\sigma^3}]$$
$$+ \sigma^{-2}\sum_{i=1}^n \mathbb{I}(y_{i,n}>0) - 3\sum_{i=1}^n \mathbb{I}(y_{i,n}>0)\sigma^{-4}(y_{i,n} - \lambda w_{i\cdot,n}Y_n - x_{i.n}\beta)^2.$$

We will show $\frac{1}{n}[\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} - \mathrm{E}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}] \xrightarrow{p} 0$ and then show $\frac{1}{n}[\frac{\partial^2 \ln L_n(\hat\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}] \xrightarrow{p} 0$.

The proof of $\frac{1}{n}[\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} - \mathrm{E}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}] \xrightarrow{p} 0$:

First, we examine some properties of $\psi(x)$ and $\{\psi(z_{i,n})\}_{i=1}^n$, where $z_{i,n} = z_{i,n}(\theta_0)$.

$$\psi'(x) = \frac{(x^2-1)\phi(x)}{\Phi(x)} + \frac{3x\phi^2(x)}{\Phi^2(x)} + \frac{2\phi^3(x)}{\Phi^3(x)}.$$

Since $\lim_{x\to-\infty}\phi(x)/[x\Phi(x)] = -1$, we have $|\psi(x)| = |-x\phi(x)/\Phi(x) - \phi^2(x)/\Phi^2(x)| \leqslant 3x^2 + C_1$ and $|\psi'(x)| \leqslant 7|x|^3 + C_2$ for some constants C_1 and C_2. Thus $\{\psi(z_{i,n})\}_{i=1}^n$ is also uniformly L_p bounded for any natural number p and $|\psi(x_1) - \psi(x_2)| \leqslant (7|x_1|^3 + 7|x_2|^3 + C_2)|x_1 - x_2|$. By Corollary 2 and Lemma A.5, we obtain that $\{\psi(z_{i,n})\}_{i=1}^n$ is also a uniformly L_2-NED random field.

Second, consider the product terms that appear in the second derivatives. Recall from Lemma 5, $\{\phi(z_{i,n})/\Phi(z_{i,n})\}_{i=1}^n$ is uniformly L_p bounded for any natural number p, and uniformly L_2-NED. Thus, by Cauchy's inequality and Lemma A.2, the terms $\{\mathbb{I}(y_{i,n}=0)\psi(z_{i,n})(w_{i\cdot,n}Y_n/\sigma)^2\}_{i=1}^n$,

$\{\mathbb{I}(y_{i,n}=0)\psi(z_{i,n})w_{i\cdot,n}Y_nx'_{i,n}\}_{i=1}^n$, $\cdots$, $\{\mathbb{I}(y_{i,n}>0)z_{i,n}^2w_{i\cdot,n}Y_n\}_{i=1}^n$ in the second derivatives of the log-likelihood function, are uniformly L_p bounded for any natural number p, and uniformly L_2-NED. Thus the WLLN applies for these NED random fields.

With the above results, to show $\frac{1}{n}[\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} - \mathrm{E}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}] \xrightarrow{p} 0$, it remains to show that $\{tr[(I_{2,n}-\lambda_0W_{22,n})^{-1}W_{22,n}]^2 - \mathrm{E}tr[(I_{2,n}-\lambda_0W_{22,n})^{-1}W_{22,n}]^2\}/n \xrightarrow{p} 0$. Notice that

$$tr[(I_{2,n}-\lambda_0W_{22,n})^{-1}W_{22,n}]^2 = tr[(I_n-\lambda_0\widetilde{W_n})^{-1}\widetilde{W_n}]^2,$$

where $\widetilde{W_n} = G_n(Y_n)W_nG_n(Y_n)$. And $\{[(I_n-\lambda_0\widetilde{W_n})^{-1}\widetilde{W_n}]_{ii}^2\}_{i=1}^n$ is a uniform NED random field from Proposition 2(2). Then the WLLN follows from the uniform boundedness of its elements:

$$\begin{aligned}
&\Big|[(I_n-\lambda G_n(Y_n)W_nG_n(Y_n))^{-1}G_n(Y_n)W_nG_n(Y_n)]_{ii}^2\Big| \\
=&\Big|\sum_{k=0}^{\infty}(1+k)\lambda^k\sum_{j_1}\cdots\sum_{j_{k+1}}w_{ij_1,n}w_{j_1j_2,n}\cdots w_{j_{k+1}i,n}\mathbb{I}(y_{i,n}>0)\mathbb{I}(y_{j_1,n}>0)\cdots\mathbb{I}(y_{j_{k+1},n}>0)\Big| \\
\leqslant&\sum_{k=0}^{\infty}(1+k)|\lambda|^k\sum_{j_1}\cdots\sum_{j_{k+1}}w_{ij_1,n}w_{j_1j_2,n}\cdots w_{j_{k+1}i,n} \\
\leqslant&\sum_{k=0}^{\infty}(1+k)|\lambda|^k||W_n||_\infty^{k+2}\leqslant\sum_{k=0}^{\infty}(1+k)\lambda_m^{-2}\zeta^{k+2}<\infty.
\end{aligned}$$

The proof of $\frac{1}{n}[\frac{\partial^2 \ln L_n(\bar\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}] \xrightarrow{p} 0$:

To show $\frac{1}{n}[\frac{\partial^2 \ln L_n(\bar\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}] \xrightarrow{p} 0$, it is sufficient to show that $\{\frac{1}{n}\frac{\partial^2 \ln L_n(\theta)}{\partial\theta\partial\theta'}\}_{i=1}^n$ is stochastically equicontinuous. This is because from Andrews (1994), $\{v_T(\cdot):T\geqslant 1\}$ is stochastically equicontinuous at τ_0, if and only if, for all sequences of random elements $\{\hat\tau_T:T\geqslant 1\}$ that satisfy $\rho(\hat\tau_T,\tau_0)\xrightarrow{p} 0$, where $\rho(\cdot,\cdot)$ is a metric, we have $v_T(\hat\tau_T)-v_T(\tau_0)\xrightarrow{p} 0$.

Most terms in the second derivatives (except one) are stochastically equicontinuous by Lemma A.7. For example, consider the first term $\frac{1}{n}\sum_{i=1}^n\mathbb{I}(y_{i,n}=0)\psi(z_{i,n})(w_{i\cdot,n}Y_n/\sigma)^2$ in $\frac{1}{n}\partial^2\ln L_n(\theta)/\partial\lambda^2$. Its stochastic equicontinuity is equivalent to that of $\frac{1}{n}\sum_{i=1}^n\psi(z_{i,n})\cdot\mathbb{I}(y_{i,n}=0)(w_{i\cdot,n}Y_n)^2$, because the parameter space of σ is compact and does not contain 0. We have shown that $|\psi(x_1)-\psi(x_2)|\leqslant(7|x_1|^3+7|x_2|^3+C_2)|x_1-x_2|$, $\{z_{i,n}(\theta)\}_{i=1}^n$ and $\{(w_{i\cdot,n}Y_n)^2\}_{i=1}^n$ are uniformly L_p bounded for any natural number p, thus the conditions of Lemma A.7 are satisfied.

The only term that Lemma A.7 does not apply to is $\frac{1}{n}tr[(I_{2,n}-\lambda W_{22,n})^{-1}W_{22,n}]^2$. However, with

$d\{\frac{1}{n}tr[(I_{2,n}-\lambda W_{22,n})^{-1}W_{22,n}]^2\}/d\lambda = \frac{2}{n}tr[(I_{2,n}-\lambda W_{22,n})^{-1}W_{22,n}]^3$, $\frac{1}{n}tr[(I_{2,n}-\lambda W_{22,n})^{-1}W_{22,n}]^2$ is stochastically equicontinuous because

$$\begin{aligned}
&|\frac{1}{n}tr[(I_{2,n}-\lambda W_{22,n})^{-1}W_{22,n}]^3| \\
\leqslant & \sup_i |\{[(I_{2,n}-\lambda W_{22,n})^{-1}W_{22,n}]^3\}_{ii}| \\
=&|(\sum_{l=0}^{\infty}\lambda^l W_{22,n}^{l+1}\sum_{l'=0}^{\infty}\lambda^{l'}W_{22,n}^{l'+1}\sum_{l''=0}^{\infty}\lambda^{l''}W_{22,n}^{l''+1})_{ii}| \\
=&|(\sum_{k=0}^{\infty}\sum_{l+l'+l''=k}\lambda^{l+l'+l''}W_{22,n}^{l+l'+l''+3})_{ii}| \\
\leqslant & \sum_{k=0}^{\infty}0.5(k+1)(k+2)\lambda_m^{-3}||\lambda_m W_{22,n}||_{\infty}^{k+3} \\
\leqslant & \sum_{k=0}^{\infty}0.5(k+1)(k+2)\lambda_m^{-3}\zeta^{k+3} < \infty.
\end{aligned}$$

□

References

[1] Abadir, K. M., Magnus, J. R. (2005). Matrix Algebra (Vol. 1). Cambridge University Press.

[2] Amaral, P., Anselin, L. (2011). Finite sample properties of Moran's I test for spatial autocorrelation in Probit and Tobit models-empirical evidence (No. 1048). GeoDa Center for Geospatial Analysis and Computation.

[3] Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. Econometrica 41, 997-1016.

[4] Amemiya, T. (1974). Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. Econometrica 42, 999-1012.

[5] Amemiya, T. (1985). Advanced Econometrics. Harvard university press.

[6] Andrews, D. W. (1992). Generic uniform convergence. Econometric theory 8, 241-257.

[7] Andrews, D. W. (1994). Asymptotics for semiparametric econometric models via stochastic equicontinuity. Econometrica 62, 43-72.

[8] Anselin, L. (1988). Spatial Econometrics: Methods and Models (Vol. 4). Springer.

[9] Anselin, L. and A.K. Bera (1998): Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics, in Handbook of Applied Economics Statistics, A. Ullah and D.E.A Giles, eds., New York: Marcel Dekker

[10] Autant-Bernard, C., LeSage, J. P. (2011). Quantifying knowledge spillovers using spatial econometric models. Journal of Regional Science 51, 471-496.

[11] Ballester, C., Calvó-Armengol, A., Zenou, Y. (2006). Who's who in networks. wanted: the key player. Econometrica, 74, 1403-1417.

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

[13] Chudik, A., Pesaran, M. H., Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. The Econometrics Journal, 14, C45-C90.

[14] Davidson, J. (1994). Stochastic Limit Theory: An Introduction for Econometricians: An Introduction for Econometricians. Oxford university press.

[15] Donfouet, H. P. P., Jeanty, P. W., Malin, É. (2012). Accounting for spatial interactions in the demand for community-based health insurance: A Bayesian spatial Tobit analysis. In 11th International Workshop Spatial Econometrics and Statistics.

[16] Flores-Lagunes, A., Schnier, K. E. (2012). Estimation of sample selection models with spatial dependence. Journal of Applied Econometrics 27, 173-204.

[17] Gallant, A. R., White, H. (1988). A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. New York: Basil Blackwell.

[18] Jenish, N. (2012). Nonparametric spatial regression under near-epoch dependence. Journal of Econometrics 167, 224-239.

[19] 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.

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

[21] 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.

[22] Kelejian, H. H., Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. International economic review 40, 509-533.

[23] Lahiri, S. N. (1996). On inconsistency of estimators based on spatial data under infill asymptotics. Sankhyā: The Indian Journal of Statistics, Series A, 403-417.

[24] Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899-1925.

[25] Lee, L. F. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics 137, 489-514.

[26] Lei, J. (2013). Smoothed spatial maximum score estimation of spatial autoregressive binary choice panel models. Working paper.

[27] LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. Geographical Analysis 32, 19-35.

[28] LeSage, J. P. and R. Pace (2009). Introduction to Spatial Econometrics. CRC Press.

[29] Marsh, T. L., Mittelhammer, R. C. (2004). Generalized maximum entropy estimation of a first order spatial autoregressive model. Advances in econometrics 18, 199-234.

[30] Olsen, R. (1978), Note on the uniqueness of the maximum likelihood estimator for the Tobit model, Econometrica 46, 1211-1215.

[31] Qu, X., Lee, L. F. (2012). LM tests for spatial correlation in spatial models with limited dependent variables. Regional Science and Urban Economics 42, 430-445.

[32] Qu, X., Lee, L. F., (2013a). Locally most powerful tests for spatial interactions in the simultaneous SAR Tobit model. Regional Science and Urban Economics 43, 307-321.

[33] Qu, X., Lee, L. F., (2013b). Estimating a Spatial Autoregressive Model with an Endogenous Spatial Weight Matrix. Working paper.

[34] Rothenberg, T. J. (1971). Identification in parametric models. Econometrica 39, 577-591.

[35] Sheldon, R. (2002). A First Course In Probability, 6/E. Pearson Education India.

[36] Shorack, G. R. (2000). Probability for Statisticians. New York: Springer.

[37] Smith, T. E. (2009). Estimation bias in spatial models with strongly connected weight matrices. Geographical Analysis, 41, 307-332.

[38] Wegge, L. L. (1974). Mean value theorem for convex functions. Journal of Mathematical Economics 1, 207-208.

[39] Xu, X., Lee, L. F. (2013). A Spatial Autoregressive Model with a Nonlinear Transformation of the Dependent Variable. Working paper.