Journal of Econometrics 188 (2015) 264–280

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Maximum likelihood estimation of a spatial autoregressive Tobit model

Xingbai Xu, Lung-fei Lee *

Department of Economics, The Ohio State University, Columbus, OH, United States

ARTICLE INFO

Article history:
Received 3 December 2013
Received in revised form
15 April 2015
Accepted 18 May 2015
Available online 9 June 2015

JEL classification:
C13
C21
C24
C63

Keywords:
Spatial autoregressive model
Tobit model
Censored data
Near-epoch dependence
Maximum likelihood

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.

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) and LeSage and Pace (2009). The two-stage least squares estimation is explored in Kelejian and Prucha (1998, 1999), while the generalized method of moments is studied in Lee (2007). The large sample properties of quasi-maximum likelihood estimation is considered in Lee (2004).

In recent years, there has been growing interest in nonlinear SAR models, as linear models cannot fully capture the characteristics of some types of data, such as censored or binary data. 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 (2015). The smoothed maximum score estimation of binary choice panel models with spatial autoregressive errors can be found in Lei (2013). Qu and Lee (2015) investigate the estimation of an SAR model with an endogenous spatial weights matrix. To study nonlinear SAR models and extend asymptotic properties of extreme estimators of nonlinear models with serial correlation (e.g. Gallant and White, 1988) to spatial correlation, some laws of large numbers (LLN) and central limit theorems (CLT) are needed. Jenish and Prucha (2009, 2012) have made fundamental contributions in this area by establishing LLN and CLT for spatial mixing and NED random fields.

In the microeconometric literature, the Tobit model has been widely studied since Tobin (1958) and 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. Some studies (e.g. Flores-Lagunes and Schnier, 2012) consider estimation or suggest tests of the spatial error Tobit model, but here we only review the literature on spatial autoregressive Tobit (SAR Tobit) models. In the literature, there are two types of SAR Tobit models (Qu and Lee, 2012): the simultaneous SAR Tobit model ($y_{i,n} = \max(0, \lambda_0 \sum_{j=1}^n w_{ij,n} y_{j,n} + x_{i,n}\beta_0 + \epsilon_{i,n})$), and the latent SAR Tobit model ($y_{i,n} = \max(0, y^*_{i,n})$, where $y^*_{i,n} = \lambda_0 \sum_{j=1}^n w_{ij,n} y^*_{j,n} + x_{i,n}\beta_0 + \epsilon_{i,n}$). The different interpretations for these two models are given in Qu and Lee (2012). So far, most studies have focused on the second type. LeSage (2000) and LeSage and Pace (2009) consider the Bayesian estimation of the latent SAR Tobit model. Marsh and Mittelhammer

* Corresponding author. Tel.: +1 614 292 5508.
E-mail addresses: xu.510@buckeyemail.osu.edu (X. Xu), lee.1777@osu.edu (L.-f. Lee).

http://dx.doi.org/10.1016/j.jeconom.2015.05.004
0304-4076/

(2004) study the performance of the generalized maximum entropy estimation of the SAR and the latent SAR Tobit models with Monte Carlo simulations. Testing of the existence of spatial correlation in the latent SAR Tobit model is carried out in Amaral and Anselin (2011) and Qu and Lee (2012). Donfouet et al. (2012) apply the latent SAR Tobit model to examine community-based health insurance using Bayesian estimation.

Compared to the latent SAR Tobit model, there are even fewer studies on the simultaneous SAR Tobit model. Autant-Bernard and LeSage (2011) consider the Bayesian estimation of this model and apply it to study knowledge spillovers. Qu and Lee (2012, 2013) examine the existence of spatial correlation in the simultaneous SAR Tobit model. To the best of our knowledge, there have been no formal studies on asymptotic properties of estimators of the simultaneous SAR Tobit model (we will call it SAR Tobit model for short, since we do not study the latent one). In this paper, we first show 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 SAR Tobit model and discuss its model coherency. In Section 3, we derive NED properties of the dependent variable and some other relevant functions of random variables. In Section 4, the identification of the SAR Tobit model and the consistency of its MLE are discussed. In Section 5, we establish the asymptotic normality of the estimator. In Section 6, we study finite sample properties and the robustness of the estimator using Monte Carlo experiments. All of the proofs for propositions and theorems are presented in the Appendices.

2. The spatial autoregressive Tobit model

The SAR Tobit model is motivated by two branches of literature in economics. One branch is concerned with peer effects from an exogenous social network. In such studies, each player chooses his/her effort, which is usually assumed to be nonnegative, and the Nash Equilibrium is exactly our model. See the detailed discussion below Eq. (2). The other branch is concerned with econometrics studies where a significant fraction of nonnegative data can be zero. Our model captures both features with a perfect information game framework in which an individual maximizes utility by choosing effort subject to a nonnegative constraint. A few empirical studies in the existing literature might appropriately be implemented by our model: (1) Rupasingha et al. (2004) investigate the environmental Kuznets curve for US counties. In the data set, facilities do not need to report to relevant environmental agents when they manufacture or process less than 25,000 pounds of a listed chemical during a year. Some counties do not have firms that individually meet these criteria. Thus, the pollutant data of these counties are censored. (2) Direct agriculture disaster payment relief for different states in US is studied in Marsh and Mittelhammer (2004), where some states do not receive any such payment (in certain years) and some spatial correlation exists. (3) LeSage (2009) examines origin–destination (OD) commuting flows from 60 districts in Toulouse, France. About 15% of the 3600 OD flows have zero values. (4) As is pointed out in Donfouet et al. (2012), community-based health insurance (CBHI) has increasing demand in rural areas in developing countries and households are more likely to pay for CBHI if other households in the same village are willing to do so. However, some households do not pay for the CBHI. (5) Spatial correlation among school district income tax rates in various school districts in Iowa is examined in Qu and Lee (2012, 2013).[1]

[1] We complete their studies with the asymptotic normality and variance of the MLE, presented in a supplementary online file (see Appendix C).

Let $\{(y_{i,n}, x_{i,n})\}_{i=1}^n$, where $y_{i,n}$ is censored such that $y_{i,n} \geqslant 0$ and $x_{i,n} \in \mathbb{R}^K$, be the sample we observe. We denote the position of individual (spatial unit) i as $s_i \in \mathbb{R}^d$, a point in the d-dimensional Euclidean space. For simplicity of notation, we also use i to represent s_i. As there are interactions among different individuals, we use an $n \times n$ matrix $W_n = (w_{ij,n})$ to represent their relative strength of direct interactions. If there is a potential 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.

$F(x) \equiv \max(0, x)$ is both a non-decreasing convex function and a Lipschitz function such that $|F(x_1) - F(x_2)| \leqslant |x_1 - x_2|$. The SAR Tobit model in this paper is specified as

$$y_{i,n} = F(\lambda_0 w_{i\cdot,n} Y_n + x_{i,n}\beta_0 + \epsilon_{i,n}), \tag{1}$$

where $w_{i\cdot,n}$ is the ith row of W_n. With generalized notations $\max(0, (x_1, \ldots, x_n)') = F((x_1, \ldots, x_n)') \equiv (\max(0, x_1), \ldots, \max(0, x_n))'$, the model can be written as

$$Y_n = \max(0, Y_n^*) = F(\lambda_0 W_n Y_n + X_n\beta_0 + \epsilon_n), \tag{2}$$

where $Y_n^* \equiv \lambda_0 W_n Y_n + X_n\beta_0 + \epsilon_n$. The model can be derived as a complete information game with each spatial unit (an agent) maximizing its linear–quadratic utility function subject to nonnegative constraints, given the actions of its links (see Ballester et al., 2006; Calvó-Armengol et al., 2009). Assume individual i's utility is $u(y_{1,n}, \ldots, y_{n,n}) = -y_{i,n}^2 + 2(\lambda_0 w_{i\cdot,n} Y_n + x_{i,n}\beta_0 + \epsilon_{i,n}) y_{i,n}$. When $\lambda_0 w_{i\cdot,n} Y_n + x_{i,n}\beta_0 + \epsilon_{i,n} > 0$, u is maximized at $y_{i,n} = \lambda_0 w_{i\cdot,n} Y_n + x_{i,n}\beta_0 + \epsilon_{i,n}$. When $\lambda_0 w_{i\cdot,n} Y_n + x_{i,n}\beta_0 + \epsilon_{i,n} \leqslant 0$, $\partial u / \partial y_{i,n} = -2y_{i,n} + 2(\lambda_0 w_{i\cdot,n} Y_n + x_{i,n}\beta_0 + \epsilon_{i,n}) < 0$ when $y_{i,n} > 0$, i.e., u is strictly decreasing when $y_{i,n} > 0$. Thus, u is maximized at $y_{i,n} = 0$. Hence, the Nash equilibrium is Eq. (2).

After a slight modification, this model can deal with the case where censoring points $c_{i,n}$'s are known and nonzero. With $c_{i,n}$ being the censoring points for i, an extended model can be $\tilde{Y}_n = \max\{C_n, \tilde{Y}_n^*\}$ where $\tilde{Y}_n^* = \lambda_0 W_n \tilde{Y}_n + X_n\beta_0 + \epsilon_n$. We can do a transformation, $Y_n = \tilde{Y}_n - C_n$, then the extended model can be rewritten as Eq. (1) with a trivial modification on the regressors, $Y_n^* = \lambda_0 Y_n + \lambda_0 W_n C_n - C_n + X_n\beta_0 + \epsilon_n$. But this paper cannot deal with the case studied in Nelson (1977) where $c_{i,n}$'s are unknown. An example of such a model is the female labor supply where $y_{i,n}$ is a market wage and $c_{i,n}$ is a reservation wage. The reservation wage can be modeled as another regression equation based on an individual's unobserved utility. We can observe market wage $y_{i,n}$ when $y_{i,n} > c_{i,n}$, but we will not observe $y_{i,n}$ while $y_{i,n} \leqslant c_{i,n}$ because those females will not work. Such a model is in the category of sample selection models. The censored model in this paper cannot handle sample selection models.

Because our model (2) is a system of nonlinear equations with censored dependent variables, it is necessary to discuss 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.

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 induced from any norm on $\mathbb{R}^d$. Assumption 1 uses the increasing domains asymptotic and rules out the scenario of infilled asymptotic.[2] This setting is introduced in Jenish and Prucha (2009, 2012) for spatial mixing and NED processes.

[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).

Assumption 2. $\zeta \equiv \lambda_m \sup_n \|W_n\|_\infty < 1$, and $\Lambda = [-\lambda_m, \lambda_m]$ is the compact parameter space of λ on the real line.

As the true parameter λ_0 must be in the parameter space, $\lambda_0 \sup_n \|W_n\|_\infty < 1$. A similar assumption for λ_0 can be found in the linear SAR model (e.g. Kelejian and Prucha, 1998). Assumption 2 is related to stability in the linear SAR model and the model coherency for the Tobit model. Amemiya (1974) discusses the model coherency for the simultaneous equation Tobit model using principal minor. From Theorem 3 in Amemiya (1974), Eq. (2) has a unique solution if and only if every principal minor of $I_n - \lambda_0 W_n$ 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 system by a contraction mapping for any possible value λ in the parameter space, as in Qu and Lee (2013). When Assumption 2 fails, it is possible that there are no or multiple solutions (see Examples 1 and 2 on p. 1006, Amemiya, 1974). It is also possible that there is a unique solution, e.g., the solution is unique for the system, $y_1 = \max(0, \lambda y_2 + a)$ and $y_2 = \max(0, b)$, no matter what the value of λ is. Even infinite solutions are possible, e.g., all $y_1 = y_2 \geqslant 0$ are solutions of the system: $y_1 = \max(0, y_2)$, $y_2 = \max(0, y_1)$. Thus Assumption 2 is a sufficient but not necessary condition for the uniqueness of the solution.

In the past literature, an incoherent model was regarded as an unsatisfactory probability model. In the recent econometric research on game estimation on discrete choices, however, researchers take an alternative view. Model incoherency reflects the presence of multiple equilibria in the model, and suggests either completing the model by specifying an equilibrium selection rule (see, e.g., Bajari et al., 2010) or set-estimation methods (see, e.g. Chernozhukov et al., 2007; Ciliberto and Tamer, 2009).

One might question whether Assumption 2 can be relaxed. When $\sup_n \|W_n\|_\infty = 1$, Assumption 2 is equivalent to $\lambda_m < 1$. When $|\lambda| > 1$, there are still at least two problems, in addition to the model coherency mentioned above. First, we need $|\lambda| < 1$ such that the Neumann series expansion $(I_n - \lambda W_n)^{-1} = I_n + \lambda W_n + \lambda^2 W_n^2 + ..$ converges. Second, for at least one type of matrix, the so-called one forward and one behind weights matrix in Koch (2012), when it is normalized, the inverses of its characteristic values are dense in $(-\infty, -1] \cup [1, \infty)$ as $n \to \infty$. Since $\ln|I_n - \lambda W_n|$ appears in the log-likelihood function, we will have $-\infty$ when λ takes a value in such a dense set. Thus, in general, without Assumption 2, the uniform convergence will be difficult to establish.

Another difficulty is to relax the compact parameter space of λ to $(-1, 1)$ if $\sup_n \|W_n\|_\infty = 1$. This is a technical problem. The compactness of the parameter space is a requirement for most consistency theorems for nonlinear estimators, such as Theorem 2.1 in Newey and McFadden (1994) and Theorem 3.3 in Gallant and White (1988); also it is often required when we try to establish the identifiable uniqueness, e.g., Lemma 4.1 in Pötscher and Prucha (1997). Consistency conditions without the compactness of parameter spaces are discussed in Section 2.6 in Newey and McFadden (1994) and on p. 108 of Amemiya (1985). From Theorem 2.7 in Newey and McFadden (1994), if the objective function is concave, then we can still obtain the consistency. But the log-likelihood function of our model does not seem to be concave, even after some parametric transformations.[4]

Finally, Assumption 2 is also closely related to the situation where the correlation between variables is weak when their distance is large. This is much more apparent for the linear SAR model with a continuous dependent variable, $Y_n = \lambda W_n Y_n + X_n\beta + \epsilon_n$, which can be regarded as a special case of the Tobit model if the censoring becomes negligible. By iteration, $Y_n = X_n\beta + \lambda W_n X_n \beta + (\lambda W_n)^2 X_n\beta + \cdots + (\lambda W_n)^m X_n \beta + \cdots + \epsilon_n + \lambda W_n \epsilon_n + \cdots$. When $\|\lambda W_n\|_\infty > 1$, more distant neighbors might have stronger impacts on a spatial unit than nearer ones. This is apparently not an interesting case, but the reason is not due to multiple Nash equilibria.

In addition to Assumption 2, we need some structures for W_n in order to establish that the dependent variable of the SAR Tobit model is an NED process. In many applications, elements in W_n are nonnegative, but in theory we can allow elements of W_n to take on both negative and positive values.[5] In addition, we assume that there are no measurement errors in W_n, as is usually done in spatial econometrics. In practice, researchers might use their knowledge to specify the elements of the weights matrix. If there are several possible specified weights matrices, one has a model selection problem and he may use some model selection procedures. Recently, there have been some attempts that treat a specified weights matrix as endogenous, and suggest estimation methods to handle the possible endogeneity issue. When the SAR model is used to study social interactions or networks, there are some attempts to study the network formation issue. However, there are distinctions between spatial networks and social networks. Social networks such as friendship networks might be randomly observed, but spatial networks are specified based on geographical or economic characteristics of spatial units instead of being randomly observed.

Assumption 3. In addition to $w_{ii,n} = 0$ for all i, the weights $w_{ij,n}$ in W_n 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 $\bar{d}_0 > 1$. That is to say, $w_{ij,n}$ can be nonzero 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 $|\lambda_0| \sum_{i=1}^n |w_{ij,n}| > \zeta$, is less than or equal to some fixed nonnegative integer that does not depend on n, denoted as N; (ii) there exists an $\alpha > d$ and a constant C_0 such that $|w_{ij,n}| \leqslant C_0/d(i,j)^\alpha$.[6]

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, one may 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 requires 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 not negligible even when their distances from i

---

[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 ith 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 conjugate $\overline{\lambda_i}$ 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] Newey and McFadden (1994) point out that it is possible to relax compactness with some nonconcave objective functions for regression models, such as a result in McDonald and Newey (1988), but their result is not applicable in our setup. Assumption D on p. 108 of Amemiya (1985) is another condition, but it is hard to verify in our model.

[5] Thanks to one referee for pointing this out.

[6] 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.

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.[7] Assumption 3(2)(ii) includes the setting of Assumption 3(1) and the case that $|w_{ij,n}|$ with an upper bound decreasing exponentially 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 $|\{j : |\lambda_0| \sum_{i=1}^n |w_{ij,n}| > \zeta\}| \leqslant N$. 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.[8] Here is an example that satisfies Assumption 3(1) but not 3(2). For simplicity, let $\lambda_0 = 1$, n be a multiple of 3 and W_n be a block diagonal matrix with each diagonal block as $\begin{bmatrix} 0 & 0 & 0 \\ 0.6 & 0 & 0 \\ 0.6 & 0 & 0 \end{bmatrix}$. Then the number of columns satisfying $\lambda_0 \sum_{i=1}^n |w_{ij,n}| > 1 > \zeta$ is $n/3$, unbounded as $n \to \infty$. However, for a symmetric W_n that satisfies Assumption 2, Assumption 3(2)(i) holds with $N = 0$, because $|\lambda| \sum_i |w_{ij,n}| = |\lambda| \sum_i |w_{ji,n}| \leqslant \zeta$.

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 1–3(2) imply that $\sup_n \|\lambda_0 W_n\|_1 < \infty$ (see the proof of Lemma 1). Furthermore, the column sum of $|W_n|^l$, where $|W_n| \equiv (|w_{ij,n}|)_{n\times n}$, multiplied by λ_0^l decays geometrically as l increases, as stated in Lemma 1.

Lemma 1. *Under Assumptions 1–3(2), $\Gamma \equiv |\lambda_0| \sup_n \|W_n\|_1 < \infty$ and $\|\lambda_0^l |W_n|^l\|_1 \leqslant \max(lN, 1)\Gamma\zeta^{l-1}$, where $|W_n| \equiv (|w_{ij,n}|)_{n\times n}$.*

This lemma has explored a feature for $|W_n|^l$ so the result is established for $\|\lambda_0^l |W_n|^l\|_1$ as a whole[9] and it will be used in the establishment of NED properties of the dependent variable. In deriving the NED, the Taylor expansion is applied to $(I_n - |\lambda_0| \cdot |W_n|)^{-1}$, and thus $|W_n|^l$ appears. This is why Lemma 1 is required.

3. Moment and NED properties of some variables

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 properties of NED random fields in Jenish and Prucha (2012) for the convenience of reference.

For any random variable υ, $\|\upsilon\|_p = [E|\upsilon|^p]^{\frac{1}{p}}$ denotes its L_p-norm. Let $Z = \{Z_{i,n}, i \in D_n, n \geqslant 1\}$ and $\upsilon = \{\upsilon_{i,n}, i \in D_n, n \geqslant 1\}$ 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 the σ-field generated by the random variables $\upsilon_{j,n}$'s with units j's located within the ball $B_i(s)$ with radius s and centered at i. Z is said to be *L_p-near-epoch dependent on υ* 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* iff $\psi(s) = O(\rho^s)$ for some $0 < \rho < 1$. The NED property is kept under summation, product (see Lemma A.2) and Lipschitz transformations.

Now we are ready to discuss the NED properties of $\{y_{i,n}\}_{i=1}^n$ and its transformations on the base $\{x_{i,n}, \epsilon_{i,n}\}_{i=1}^n$. That is to say, $\mathcal{F}_{i,n}(s) \equiv \sigma(\{x_{j,n}, \epsilon_{j,n} : d(i,j) \leqslant s\})$. And the conditions for $x_{i,n}$ and $\epsilon_{i,n}$ are summarized in the following assumption:

Assumption 4. $\sup_{1\leqslant k\leqslant K,i,n} \mathrm{E}|x_{ik,n}|^2 < \infty$; $\sup_{i,n} \mathrm{E}|\epsilon_{i,n}|^2 < \infty$.

Notice that we do not impose normality on the disturbances in Assumption 4 to highlight some general stochastic structures implied solely by the model (1) without a distributional assumption. In later analyses, we will often deal with $\{w_{i\cdot,n}Y_n\}_{i=1}^n$, $\{y_{i,n}^* \equiv \lambda_0 w_{i\cdot,n}Y_n + x_{i,n}\beta_0 + \epsilon_{i,n}\}_{i=1}^n$ and $\{z_{i,n}(\theta) \equiv (y_{i,n} - \lambda w_{i\cdot,n}Y_n - x_{i,n}\beta)/\sigma\}_{i=1}^n$. We summarize their moment and NED properties in the following proposition. Recall that α in the following proposition is a parameter in Assumption 3(2)(ii): $|w_{ij,n}| \leqslant C_0/d(i,j)^\alpha$.

Proposition 1. (1) *Under Assumption 2, if $\sup_{1\leqslant k\leqslant K,i,n} \mathrm{E}|x_{ik,n}|^p < \infty$ and $\sup_{i,n} \mathrm{E}|\epsilon_{i,n}|^p < \infty$ for some $p \geqslant 1$, then $\{y_{i,n}\}_{i=1}^n$, $\{w_{i\cdot,n}Y_n\}_{i=1}^n$, $\{z_{i,n}(\theta)\}_{i=1}^n$ and $\{y_{i,n}^*\}_{i=1}^n$ are all uniformly L_p bounded.*

(2) *Under Assumptions 1–3(1) and 4, $\{y_{i,n}\}_{i=1}^n$ is geometrically L_2-NED on $\{x_{i,n}, \epsilon_{i,n}\}_{i=1}^n$: $\|y_{i,n} - E[y_{i,n}|\mathcal{F}_{i,n}(s)]\|_2 \leqslant C(\zeta^{1/\bar{d}_0})^s$ for some $C > 0$ that does not depend on i and n. The same conclusion also holds for $\{w_{i\cdot,n}Y_n\}_{i=1}^n$, $\{z_{i,n}(\theta)\}_{i=1}^n$ and $\{y_{i,n}^*\}_{i=1}^n$.*

(3) *Under Assumptions 1–3(2) and 4, $\{y_{i,n}\}_{i=1}^n$ is L_2-NED on $\{x_{i,n}, \epsilon_{i,n}\}_{i=1}^n$: $\|y_{i,n} - E[y_{i,n}|\mathcal{F}_{i,n}(s)]\|_2 \leqslant C/s^{\alpha-d}$ for some $C > 0$ that does not depend on i and n. The same conclusion also holds for $\{w_{i\cdot,n}Y_n\}_{i=1}^n$, $\{z_{i,n}(\theta)\}_{i=1}^n$ and $\{y_{i,n}^*\}_{i=1}^n$.*

As shown below, we not only deal with linear functions of the dependent variables, but also with some of their nonlinear functions. To analyze the uniform L_p boundedness of these nonlinear functions, we can bound them and their derivatives by some polynomial functions. With these polynomial bounds and Proposition 1(1), we can establish the uniform L_p boundedness. With Lemma A.1, which relates relevant polynomial functions to the L_2-NED property of those of basic dependent variables, we obtain Lemma A.4. Lemma A.4, which is a generalization beyond Lipschitz functions to a wider range of functions, can preserve the NED property of their arguments. Since the disturbances are normally distributed in the Tobit model, we will deal with the distribution and density functions of the standard normal distribution quite often. 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. The NED properties of relevant functions of the model are summarized in Lemma A.9.

Another nonlinear transformation of $y_{i,n}$ in the log-likelihood function is the dichotomous indicator $\mathbb{I}(y_{i,n} > 0)$. 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 $\{y_{i,n}^*\}_{i=1}^n$.

Proposition 2. (1) *Under Assumptions 1–3(1) and 4, 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 1–3(2) and 4, 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 uniformly L_2-NED 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 $C > 0$.*

[7] 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).

[8] 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 usually generates weak dependence, as noted in Chudik et al. (2011).

[9] The inequality may not be valid for $(\lambda_0\|W_n\|_1)^l$.

Because the joint mixed probability density function of $\{y_{1,n}, \ldots, y_{n,n}\}$ 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 Lemmas A.7 and A.8 and apply them to the model. Let $\widetilde{W}_n = G_n(Y_n)W_nG_n(Y_n)$, where $G_n(Y_n) = diag(\mathbb{I}(y_{1,n} > 0), \ldots, \mathbb{I}(y_{n,n} > 0))$, and $r_{i,n}(\lambda) = ((I_n - \lambda\widetilde{W}_n)^{-1}\widetilde{W}_n)_{ii}$. With Proposition 2 and $\eta = \zeta^{1/3d_0}$, Lemmas A.7 and A.8 imply the following results:

Proposition 3. (1) *Under Assumptions 1–3(1) and 4, if the essential supremums of densities of* $\{y^*_{i,n}\}_{i=1}^n$ *are uniformly bounded in i and n, then both* $\{r_{i,n}(\lambda_0)\}_{i=1}^n$ *and* $\{(([I_n - \lambda_0\widetilde{W}_n)^{-1}\widetilde{W}_n]^2)_{ii}\}_{i=1}^n$ *are uniformly and geometrically L_2-NED with coefficients* $s(\zeta^{1/3\bar{d}_0})^s$ *and* $s^2(\zeta^{1/3\bar{d}_0})^s$ *respectively.*

(2) *Under Assumptions 1–3(2) and 4, if the essential supremums of densities of* $\{y^*_{i,n}\}_{i=1}^n$ *are uniformly bounded in i and n, then* $\{r_{i,n}(\lambda_0)\}_{i=1}^n$ *and* $\{(([I_n - \lambda_0\widetilde{W}_n)^{-1}\widetilde{W}_n]^2)_{ii}\}_{i=1}^n$ *are uniformly L_2-NED random fields with coefficient* $1/s^{(\alpha-d)/3}$.

From Theorem 1 in Jenish and Prucha (2012), for a uniform L_1-NED random field $\{Z_{i,n}\}_{i=1}^n$ on some suitable α-mixing random field with $\sup_{i,n}\|Z_{i,n}\|_p < \infty$ for some $p > 1$, $\frac{1}{n}\sum_{i=1}^n(Z_{i,n} - \mathrm{E}Z_{i,n}) \xrightarrow{L_1} 0$. Since a uniform L_2-NED random field is also uniformly L_1-NED, the weak LLN also holds. The CLT for NED random fields requires more conditions. Assume $Z = \{Z_{i,n}\}_{i=1}^n$ is a zero-mean uniform L_2-NED random field on some suitable α-mixing random field. If (1) Z is uniformly $L_{2+\delta}$ integrable for some $\delta > 0$, (2) $\inf_n \frac{1}{n}\sigma_n^2 > 0$, where $\sigma_n^2 = \mathrm{Var}(\sum_{i=1}^n Z_{i,n})$ and (3) NED coefficients satisfy $\sum_{r=1}^\infty r^{d-1}\psi(r) < \infty$, then $\sigma_n^{-1}\sum_{i=1}^n Z_{i,n} \xrightarrow{d} N(0, 1)$. Thus, with the uniform L_p boundedness and uniform L_2-NED properties established 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 ML to estimate the true parameters. The Tobit model is established under the distributional specification that the error terms are normally distributed.

Assumption 5. For each n, $\epsilon_{i,n}$'s are *i.i.d.* $N(0, \sigma^2)$ random variables; X_n and ϵ_n are independent.

Assumption 6. (i) $\{x_{i,n}\}_{i=1}^n$ is an α-mixing random field with α-mixing coefficient $\alpha(u, v, r) \leqslant (u+v)^\tau\hat{\alpha}(r)$ for some $\tau \geqslant 0$, where $\hat{\alpha}(r)$ satisfies $\sum_{r=1}^\infty r^{d-1}\hat{\alpha}(r) < \infty$. (ii) $\sup_{i,k,n}\|x_{ik,n}\|_{4+\Delta} < \infty$ for some $\Delta > 0$.

Assumption 7. The parameter space Θ of $\theta = (\lambda, \beta', \sigma)'$ is a compact subset of $\mathbb{R}^{K+2}$.

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

$$\begin{aligned}
\ln L_n(\theta) &= \sum_{i=1}^n \mathbb{I}(y_{i,n} = 0)\ln\left[1 - \Phi\left(\frac{\lambda}{\sigma}w_{i\cdot,n}Y_n + x_{i,n}\frac{\beta}{\sigma}\right)\right] \\
&\quad - \frac{1}{2}\ln(2\pi\sigma^2)\sum_{i=1}^n \mathbb{I}(y_{i,n} > 0) \\
&\quad + \ln\det(I_{2,n} - \lambda W_{22,n}) - \frac{1}{2}\sum_{i=1}^n \mathbb{I}(y_{i,n} > 0) \\
&\quad \times \left(\frac{1}{\sigma}y_{i,n} - \frac{\lambda}{\sigma}w_{i\cdot,n}Y_n - x_{i,n}\frac{\beta}{\sigma}\right)^2 \\
&= \sum_{i=1}^n \mathbb{I}(y_{i,n} = 0)\ln\Phi(z_{i,n}(\theta)) \\
&\quad - \frac{1}{2}\ln(2\pi\sigma^2)\sum_{i=1}^n \mathbb{I}(y_{i,n} > 0) \\
&\quad + \ln\det(I_{2,n} - \lambda W_{22,n}) - \frac{1}{2}\sum_{i=1}^n \mathbb{I}(y_{i,n} > 0)z_{i,n}^2(\theta), \quad (3)
\end{aligned}$$

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 $W_{22,n}$, $\det(A)$ is the absolute value of the determinant of the matrix A, 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 $I_{2,n}$ and the positions of elements of $W_{22,n}$ in W_n are stochastic, 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 $\widetilde{W}_n = G_n(Y_n)W_nG_n(Y_n)$. Integrating both sides of $d\ln\det(I_{2,n} - \lambda W_{22,n})/d\lambda = -tr[(I_{2,n} - \lambda W_{22,n})^{-1}W_{22,n}] = -tr(\sum_{l=0}^\infty \lambda^l W_{22,n}^{l+1})$, where the first equality is from a rule for matrix differentiation (Amemiya, 1985, p. 461), we obtain

$$\begin{aligned}
\ln\det(I_{2,n} - \lambda W_{22,n}) &= -\sum_{l=1}^\infty (\lambda^l/l)tr(W_{22,n}^l) \\
&= -\sum_{l=1}^\infty (\lambda^l/l)tr(\widetilde{W}_n^l) = \ln\det(I_n - \lambda\widetilde{W}_n), \quad (4)
\end{aligned}$$

where the second equality holds because $trW_{22,n}^l = tr\widetilde{W}_n^l$.

Under Assumptions 5 and 6, by Proposition 1 and Lemma A.9, $\{y_{i,n}\}_{i=1}^n$, $\{y^*_{i,n}\}_{i=1}^n$, $\{z_{i,n}(\theta)\}_{i=1}^n$ and $\{\frac{\phi(z_{i,n}(\theta))}{\Phi(z_{i,n}(\theta))}\}_{i=1}^n$ are all uniformly $L_{4+\Delta}$ bounded, while $\{z_{i,n}(\theta)^2\}_{i=1}^n$, $\{\ln\Phi(z_{i,n}(\theta))\}_{i=1}^n$ and $\{\phi(z_{i,n}(\theta))z_{i,n}(\theta)/\Phi(z_{i,n}(\theta))\}_{i=1}^n$ are uniformly $L_{2+\Delta/2}$ bounded. In the following lemma, we show that the additional boundedness condition for $\{y^*_{i,n}\}_{i=1}^n$ in Proposition 2 is also satisfied. Consequently, $\{\mathbb{I}(y_{i,n} > 0)\}_{i=1}^n$ is a uniform L_2-NED random field.

Lemma 2. *Under Assumptions 1–3 and 5, 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 with a finite sample, 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 4. *Under Assumptions 2 and 5, if* $X_n'X_n$ *is invertible with probability 1,* $W_n + W_n' \neq 0$, *and there exists* $j \neq j'$ *such that* $\sum_{i=1}^n w_{ij,n}^2 \neq \sum_{i=1}^n w_{ij',n}^2$, *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 proof is in the Appendices.

Theorem 1. *Under Assumptions 1–8, the MLE of model (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 $z_{i,n}(\theta) \equiv (y_{i,n} - \lambda w_{i\cdot,n}Y_n - x_{i,n}\beta)/\sigma$.

$$\begin{aligned}\frac{\partial \ln L_n(\theta)}{\partial \lambda} &= -\sum_{i=1}^{n} \mathbb{I}(y_{i,n}=0)\frac{\phi(z_{i,n}(\theta))w_{i\cdot,n}Y_n}{\sigma\Phi(z_{i,n}(\theta))} \\ &\quad - tr[(I_{2,n}-\lambda W_{22,n})^{-1}W_{22,n}] \\ &\quad + \sum_{i=1}^{n}\mathbb{I}(y_{i,n}>0)\sigma^{-2} \\ &\quad \times (y_{i,n}-\lambda w_{i\cdot,n}Y_n - x_{i.n}\beta)w_{i\cdot,n}Y_n, \end{aligned} \tag{5}$$

$$\begin{aligned}\frac{\partial \ln L_n(\theta)}{\partial \beta} &= -\sum_{i=1}^{n} \mathbb{I}(y_{i,n}=0)\frac{\phi(z_{i,n}(\theta))x_{i,n}'}{\sigma\Phi(z_{i,n}(\theta))} \\ &\quad + \sum_{i=1}^{n}\mathbb{I}(y_{i,n}>0)\sigma^{-1}z_{i,n}(\theta)x_{i,n}', \end{aligned} \tag{6}$$

$$\begin{aligned}\frac{\partial \ln L_n(\theta)}{\partial \sigma} &= \sum_{i=1}^{n} \mathbb{I}(y_{i,n}=0)\sigma^{-2}\frac{\phi(z_{i,n}(\theta))(\lambda w_{i\cdot,n}Y_n + x_{i.n}\beta)}{\Phi(z_{i,n}(\theta))} \\ &\quad - \frac{1}{\sigma}\sum_{i=1}^{n}\mathbb{I}(y_{i,n}>0) \\ &\quad + \sum_{i=1}^{n}\mathbb{I}(y_{i,n}>0)\sigma^{-3}(y_{i,n}-\lambda w_{i\cdot,n}Y_n - x_{i.n}\beta)^2. \end{aligned} \tag{7}$$

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

$$q_{i,n}(\theta) = \begin{pmatrix} -\mathbb{I}(y_{i,n}=0)\dfrac{\phi(z_{i,n}(\theta))w_{i\cdot,n}Y_n}{\sigma\Phi(z_{i,n}(\theta))} + \mathbb{I}(y_{i,n}>0)\dfrac{z_{i,n}(\theta)w_{i\cdot,n}Y_n}{\sigma} - r_{i,n}(\lambda) \\ -\mathbb{I}(y_{i,n}=0)\dfrac{\phi(z_{i,n}(\theta))x_{i,n}'}{\sigma\Phi(z_{i,n}(\theta))} + \mathbb{I}(y_{i,n}>0)\dfrac{z_{i,n}(\theta)x_{i,n}'}{\sigma} \\ \mathbb{I}(y_{i,n}=0)\dfrac{\phi(z_{i,n}(\theta))(\lambda w_{i\cdot,n}Y_n + x_{i.n}\beta)}{\Phi(z_{i,n}(\theta))\sigma^2} - \dfrac{1}{\sigma}\mathbb{I}(y_{i,n}>0) + \mathbb{I}(y_{i,n}>0)\dfrac{z_{i,n}(\theta)^2}{\sigma} \end{pmatrix}. \tag{8}$$

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

Assumption 9. θ_0 is in the interior of Θ.

Assumption 10. (i) $\sup_{i,k,n}\|x_{ik,n}\|_{8+\delta} < \infty$ for some $\delta > 0$. (*ii*) For some $0 < \tilde{\delta} < 2 + \delta/2$, the α-mixing coefficient of $\{x_{i,n}\}_{i=1}^n$ in Assumption 6 satisfies $\sum_{r=1}^{\infty} r^{d(\tau_*+1)-1}\hat{\alpha}(r)^{\tilde{\delta}/(4+2\tilde{\delta})} < \infty$, where $\tau_* = \tilde{\delta}\tau/(2+\tilde{\delta})$.

Assumption 11. $\Sigma_0 = \lim_{n\to\infty}\Sigma_n$ exists and is nonsingular, where $\Sigma_n = \frac{1}{n}\mathrm{Var}\sum_{i=1}^n q_{i,n}(\theta_0)$.

In Assumption 3(ii), we assume $0 \leqslant |w_{ij,n}| \leqslant C_0/d(i,j)^{\alpha}$. But a faster decreasing rate is required to obtain asymptotic normality.

Assumption 12. $\alpha > d \cdot \max(7 + 24\delta^{-1},\ 5 + 32\delta^{-1} + 64\delta^{-2})$.

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 12 is not needed. Under Assumption 3(2)(ii), Assumption 12 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 of $\max[s^{-(\alpha-d)/3},\ s^{-(\alpha-d)\delta/(8+2\delta)}]^{\delta/(8+2\delta)}$. To satisfy the condition $\sum_{s=1}^{\infty} s^{d-1}\max[s^{-(\alpha-d)/3},\ s^{-(\alpha-d)\delta/(8+2\delta)}]^{\delta/(8+2\delta)} < \infty$, we thus need Assumption 12. If $\{x_{i,n}\}$ has Gaussian or exponential decreasing tails or is uniformly bounded, then δ can be arbitrarily large. For such cases, as a consequence, Assumption 12 becomes $\alpha > 7d$.

Proposition 5. *In addition to Assumptions 1, 2 and 5–11, suppose either Assumption 3(1), or Assumptions 3(2) and 12 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 and 5–11, if either Assumption 3(1), or Assumptions 3(2) and 11 hold, then the MLE of model (2) has* $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma_0^{-1})$.

6. Monte Carlo simulations

In this section, we perform some Monte Carlo studies to investigate the finite sample performance of the MLE and the robustness of estimates under non-normal disturbances. Our simulation studies are based on some characteristics of the empirical example in Qu and Lee (2012, 2013) that examines tax competition among local governments.[10] Specifically, they study spatial effects when local governments in Iowa set their school district income tax rates. In Iowa, this type of surtax ranges from 0% to 20%. In 2009, 18.3% of the school districts out of 361 in Iowa had 0% tax rates. Thus, the SAR Tobit model is a suitable model for this example. The theoretical background, detailed descriptive statistics of the data, and the data source can be found in Qu and Lee (2012). We choose exogenous variables with possibly significant marginal effects from their model, $x = (1, x_2, x_3, x_4, x_5)$, where $x_2, \ldots, x_5$ represent the average income (in \$1000), the percentage of white students, pupil/taxpayer (%) and property rates, respectively. We also let the true parameters in the simulation be close to the estimator in the empirical example: $\lambda_0 = 0.2$, $\beta_0' = (12, -0.4, 0.1, 1, -0.5)$ and $\sigma_0 = 5$. Thus, we consider the model $Y_n = F(\lambda W_n Y_n + X_n\beta + \epsilon_n)$, where $n = 361$ and $\epsilon_{i,n}$'s are i.i.d. $N(0, \sigma_0^2)$. The spatial weights matrix W_n is row-normalized, representing the contiguity relationship between different school districts. Specifically, W_n is row-normalized from W_n^*, where $w_{ij,n}^* = 1$ if two different school districts i and j share some common borders, otherwise 0, and $w_{ii,n}^* = 0$.

To see more MC results with variations on the strength of spatial interactions and sample size, we also try $\lambda_0 = 0.5$ and another sample size, 1121, in the experiments. When $n = 1121$, W_n is a block diagonal matrix with two blocks. The first block is the 361×361 matrix when $n = 361$, as discussed in the above paragraph, and the second block is a 760×760 matrix representing the contiguity relationship of 760 counties in the 10 Upper Great Plains States.[11] W_n is also row-normalized. The first 361 $x_{i,n}$'s are the same as those in the empirical example, and the last 760 are random samples from a multivariate normal distribution with its mean and covariance matrix estimated from the first 361.

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 $\|Y_n - F(\lambda_0 W_n Y_n +$

[10] Thank a co-editor and the associate editor of this journal for their suggestion in designing the simulation based on empirical examples so that the MCs can capture realistic features in its design.

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

Table 1
Estimation results.

n	true	λ_0 0.2	β_{10} 12	β_{20} −0.4	β_{30} 0.1	β_{40} 1	β_{50} −0.5	σ_0 5
361	mean	0.1803	12.1675	−0.4055	0.1025	0.9643	−0.5019	4.9668
	std	0.0958	4.1465	0.0576	0.0307	0.4796	0.1301	0.2254
	med	0.1814	12.1655	−0.4039	0.1017	0.9682	−0.5017	4.9636
	$q_{0.25}$	0.1179	9.5477	−0.4445	0.0824	0.6533	−0.5899	4.8181
	$q_{0.75}$	0.2480	15.0297	−0.3686	0.1221	1.2942	−0.4133	5.1111
1121	mean	0.1946	11.9754	−0.4002	0.1007	0.9989	−0.5011	4.9880
	std	0.0575	2.2090	0.0271	0.0174	0.2783	0.0773	0.1322
	med	0.1949	11.9789	−0.3997	0.1003	0.9961	−0.5019	4.9869
	$q_{0.25}$	0.1569	10.3960	−0.4180	0.0887	0.8147	−0.5501	4.9011
	$q_{0.75}$	0.2341	13.4676	−0.3827	0.1124	1.1899	−0.4458	5.0752

n	true	λ_0 0.5	β_{10} 12	β_{20} −0.4	β_{30} 0.1	β_{40} 1	β_{10} −0.5	σ_0 5
361	mean	0.4828	12.2535	−0.4071	0.1024	0.9669	−0.5022	4.9729
	std	0.0697	4.0974	0.0564	0.0300	0.4696	0.1278	0.2118
	med	0.4844	12.2511	−0.4047	0.1014	0.9674	−0.5027	4.9677
	$q_{0.25}$	0.4398	9.7218	−0.4427	0.0824	0.6579	−0.5883	4.8315
	$q_{0.75}$	0.5322	14.9077	−0.3710	0.1210	1.2829	−0.4142	5.1091
1121	mean	0.4949	11.9888	−0.4004	0.1008	0.9995	−0.5004	4.9901
	std	0.0416	2.1851	0.0263	0.0170	0.2726	0.0759	0.1203
	med	0.4960	12.0760	−0.4002	0.1006	1.0085	−0.5001	4.9889
	$q_{0.25}$	0.4668	10.4038	−0.4168	0.0892	0.8178	−0.5498	4.9167
	$q_{0.75}$	0.5222	13.4577	−0.3831	0.1124	1.1770	−0.4474	5.0647

$\epsilon_{i,n}$ *iid* $\sim N(0, \sigma_0^2)$. Repetition: 1000. $q_{0.25}$: The 25% quantile, $q_{0.75}$: The 75% quantile.

$X_n\beta_0 + \epsilon_n)\|_\infty < 10^{-6}$. We obtain the empirical mean and standard deviation, as well as median and quantiles of the estimates based on 1000 replications for each of the experiments. For each experiment, W_n and X_n are fixed for all the 1000 replications.

The estimation results are summarized in Table 1. The biases of the estimates are rather small for both sample sizes 361 and 1121. The medians are close to their true values. As the sample size n increases, the standard errors of the estimates decrease, so do their ranges constructed with 25 and 75 percentiles. The simulation results are compatible with the theoretical implication of Theorem 1.

In addition, to investigate the sensitivity of estimates on the distribution of disturbances, we try two different non-normal distributions, namely a uniform distribution $\sigma_0 U(-\sqrt{3}, \sqrt{3})$, and a mixed normal distribution (with half probability $\sigma_0 N(4/\sqrt{17}, 1/17)$, half probability $\sigma_0 N(-4/\sqrt{17}, 1/17)$). Simple calculations show that both distributions have the same standard deviation σ_0. Notice that the density of this mixed normal distribution has two peaks. As can be seen in Table 2, when the disturbance terms are uniformly distributed, biases for all $\hat{\beta}_k$'s increase as n increases from 361 to 1121. And under both distributions, biases for $\hat{\sigma}$ are large, and they are kept almost unchanged when n increases from 361 to 1121. The evidence strongly indicates the inconsistency of our estimator under misspecification of the disturbance terms.[12]

7. Conclusion

After describing some economic applications of the SAR Tobit model, this paper examines the ML estimation for this model. We establish the NED properties of the dependent variable and some relevant functions of the dependent variable. With the LLN and CLT of the NED random field, we establish the consistency and the asymptotic normality of the MLE of this model. Monte Carlo simulations, based on the empirical study of the school district income tax rate in Iowa in Qu and Lee (2012, 2013), show that biases and standard deviations of MLE decrease as sample sizes increase. And medians of estimates are also increasingly close to their true values. As the MLE method is computationally tractable, we can use it to estimate an SAR Tobit model for empirical studies of samples involving spatial correlation and censored data.

This paper focuses only on the large sample properties of the MLE of the SAR Tobit model. And it does have a few limitations. Most notably, we only consider the case in which the error terms are independently, identically and normally distributed. Nevertheless, this paper has developed a solid foundation for future studies: (1) The estimation of the SAR Tobit model with multiple weights matrices could be considered. Because the ML approach would not be computationally tractable for models with multiple weights matrices with samples of large sizes, one needs to develop computationally tractable methods for its estimation. (2) The spatial error Tobit model could be considered. For such a model, the issue is how to develop efficient estimation methods. In such a case, multivariate joint probability for censored observations will be involved, and hence the simulation estimation method is a possible approach, but asymptotic properties of such possible estimators due to simulation need to be understood. (3) It would also be interesting and important to develop distributional-free estimation approaches for the SAR Tobit model.

Acknowledgments

An earlier version of this paper was presented in the Econometric seminar at OSU, the New York Camp Econometrics 2014, and the China Meeting of the Econometric Society in Xiamen 2014. We thank the audience for their comments. We also would like to acknowledge valuable discussions and suggestions we received from Robert DeJong, Nazgul Jenish and Xi Qu. We thank a co-editor, an associate editor and two referees of this journal for their suggestions on improving the presentation of this paper.

Appendix A. Some useful lemmas

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

Lemma A.1 (*Lemma 17.15 in Davidson, 1994*)**.** *Let B and ρ be two nonnegative 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. *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, $\|[(A+B)^l - B^l]e\|_\infty \leqslant |A|_{\max}\sum_{h=0}^{l-1}\|B\|_\infty^h\|A+B\|_1^{l-1-h}\|e\|_1$.*

[12] For readers who may be interested, a supplemental file provides additional simulation results under various data generating processes besides those presented in the main text (see Appendix C).

Table 2
Robust check of estimates.

distribution	n	true	λ_0 0.2	β_{10} 12	β_{20} −0.4	β_{30} 0.1	β_{40} 1	β_{50} −0.5	σ_0 5
uniform	361	mean	0.1740	12.0245	−0.4073	0.1000	1.0049	−0.4883	5.2737
		std	0.0970	4.2987	0.0569	0.0321	0.5327	0.1420	0.1890
		med	0.1775	11.7996	−0.4050	0.0999	1.0339	−0.4893	5.2758
		$q_{0.25}$	0.1102	8.9364	−0.4433	0.0783	0.6656	−0.5798	5.1511
		$q_{0.75}$	0.2397	15.1041	−0.3691	0.1219	1.3511	−0.3943	5.3969
	1121	mean	0.1919	11.7782	−0.3928	0.0977	0.9787	−0.4878	5.2578
		std	0.0606	2.2492	0.0266	0.0169	0.2916	0.0787	0.1117
		med	0.1933	11.7711	−0.3926	0.0981	0.9892	−0.4876	5.2573
		$q_{0.25}$	0.1520	10.1608	−0.4115	0.0859	0.7845	−0.5442	5.1780
		$q_{0.75}$	0.2343	13.3425	−0.3746	0.1087	1.1730	−0.4332	5.3315
mixed normal	361	mean	0.1801	12.4155	−0.4130	0.0944	1.0931	−0.4972	5.5238
		std	0.1011	4.5158	0.0570	0.0325	0.5537	0.1508	0.1590
		med	0.1842	12.4687	−0.4117	0.0932	1.1020	−0.4989	5.5217
		$q_{0.25}$	0.1169	9.5443	−0.4497	0.0724	0.7249	−0.5969	5.4137
		$q_{0.75}$	0.2487	15.3543	−0.3738	0.1155	1.4563	−0.4032	5.6328
	1121	mean	0.1929	11.6609	−0.3920	0.0961	1.0058	−0.4875	5.5244
		std	0.0598	2.3620	0.0272	0.0180	0.3066	0.0852	0.0883
		med	0.1924	11.6489	−0.3909	0.0965	1.0079	−0.4915	5.5227
		$q_{0.25}$	0.1546	10.0934	−0.4102	0.0837	0.8002	−0.5442	5.4650
		$q_{0.75}$	0.2334	13.2495	−0.3735	0.1084	1.2119	−0.4276	5.5874

$\epsilon_{i,n}$ *iid* $\sim (0, \sigma_0^2)$. Repetition: 1000. $q_{0.25}$: The 25% quantile, $q_{0.75}$: The 75% quantile, uniform distribution: $\sigma_0 U(-\sqrt{3}, \sqrt{3})$, mixed normal distribution: half probability $\sigma_0 N(4/\sqrt{17}, 1/17)$, half probability $\sigma_0 N(-4/\sqrt{17}, 1/17)$.

Proof of Lemma A.3. 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, $\|Me\|_\infty \leqslant |M|_{\max}\|e\|_1$. Thus, for any $h = 0, \ldots, l-1$, $\|B^h A W^{l-1-h} e\|_\infty \leqslant \|B^h\|_\infty \|AW^{l-1-h}e\|_\infty \leqslant \|B^h\|_\infty |A|_{\max} \|W^{l-1-h}e\|_1 \leqslant |A|_{\max}\|B^h\|_\infty \|W\|_1^{l-1-h}\|e\|_1$. Together, we have the result. □

Lemma A.4. *$G(x)$: 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.4. Because $p/(a+1) > 2$, with Lemma A.1, we have

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

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 $p \geqslant 1$.[13] 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.5. *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.

(2) If $\{f_n(u) : R^p \to R\}_{i=1}^n$ is stochastically equicontinuous (SE) with respect to u, then $\{f_n(u(v)) : R^p \to R\}_{i=1}^n$ is SE 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 $\|u_1 - u_2\| < \delta$, $|f_n(u_1) - f_n(u_2)| < \epsilon$ for any n. When $\|v_1 - v_2\| < \delta/C$, $\|u(v_1) - u(v_2)\| \leqslant \delta$, 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 SE, 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 $\|v_1 - v_2\| < \delta/C$, $\|u(v_1) - u(v_2)\| \leqslant \delta$, 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.6. *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 SE with respect to θ, where the parameter space $\Theta(\subset R^K)$ of θ is bounded.*

Proof.

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

where the third inequality comes from C_r-inequality. By Cauchy's inequality,[14]

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

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

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

Hence, by Markov's inequality, $\frac{1}{n}\sum_{i=1}^{n}\{2K^{a-1}\sum_{j=1}^{K}|x_{ij,n}|^{a}(\sup_{\theta\in\Theta}|\theta_j|)^{a}+1\}|h(x_{i,n})x_{ik,n}|=O_p(1)$. Thus, $\{\frac{1}{n}\sum_{i=1}^{n}f(x_{i,n}\theta)h(x_{i,n})\}_{n=1}^{\infty}$ is SE with respect to θ by Lemma 1(a) in Andrews (1992). □

Lemma A.7. *Under Assumption 1, let $A_n=(a_{ij,n})$ be an $n\times n$ nonstochastic matrix satisfying $a_{ij,n}=0$ 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 $G_n=diag(v_{1,n},\ldots,v_{n,n})$. Then, for any positive integer l,*

(i) $\{g_{i,ln}\equiv(G_nA_nG_n)_{ii}^{l}\}_{i=1}^{n}$ 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) $\{u_{i,n}\equiv[(I_n-G_nA_nG_n)^{-1}G_nA_nG_n]_{ii}^{l}\}_{i=1}^{n}$ 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 $\overline{C_l}>0$.

Proof of Lemma A.7. (i) $(G_nA_nG_n)_{ii}^{l}=\sum_{j_1}\cdots\sum_{j_{l-1}}a_{ij_1,n}a_{j_1j_2,n}\cdots a_{j_{l-1}i,n}v_{i,n}^2v_{j_1,n}^2\cdots v_{j_{l-1},n}^2$. 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 $m>l$,

$$\left\|v_{i,n}^2v_{j_1,n}^2\cdots v_{j_{l-1},n}^2-\mathrm{E}\left[v_{i,n}^2v_{j_1,n}^2\cdots v_{j_{l-1},n}^2|\mathcal{F}_{i,n}(m\bar{d_0})\right]\right\|_2$$
$$\leqslant\sum_{h=0}^{l-1}\|v_{j_h,n}^2-\mathrm{E}[v_{j_h,n}^2|\mathcal{F}_{i,n}(m\bar{d_0})]\|_2$$
$$\leqslant 2\sum_{h=0}^{l-1}\|v_{j_h,n}-\mathrm{E}[v_{j_h,n}|\mathcal{F}_{i,n}(m\bar{d_0})]\|_2$$
$$\leqslant 2\sum_{h=0}^{l-1}\|v_{j_h,n}-\mathrm{E}[v_{j_h,n}|\mathcal{F}_{j_h,n}((m-h)\bar{d_0})]\|_2\leqslant 2\sum_{h=0}^{l-1}C\eta^{m-h},$$

where the second inequality follows from that v^2 is 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,

$$\|g_{i,ln}-\mathrm{E}[g_{i,ln}|\mathcal{F}_{i,n}(m\bar{d_0})]\|_2$$
$$\leqslant\sum_{j_1}\cdots\sum_{j_{l-1}}|a_{ij_1,n}a_{j_1j_2,n}\cdots a_{j_{l-1}i,n}| \cdot\|v_{i,n}^2v_{j_1,n}^2\cdots v_{j_{l-1},n}^2-\mathrm{E}[v_{i,n}^2v_{j_1,n}^2\cdots v_{j_{l-1},n}^2|\mathcal{F}_{i,n}(m\bar{d_0})]\|_2$$
$$\leqslant\|A_n\|_\infty^l 2\max(1,C)\sum_{h=0}^{l-1}\eta^{m-h}\leqslant\left(2\max(1,C)\sum_{h=0}^{l-1}\eta^{l-h}\right)\eta^m$$
$$<(2\eta\max(1,C)/(1-\eta))\eta^m.$$

(ii) Notice

$$[(I_n-G_nA_nG_n)^{-1}G_nA_nG_n]_{ii}^{l}=\left(\prod_{j=1}^{l}\sum_{L_j=1}^{\infty}[G_nA_nG_n]^{L_j}\right)_{ii}$$
$$=\sum_{k=l}^{\infty}\sum_{L_1+L_2+\cdots+L_l=k}[G_nA_nG_n]_{ii}^{k}=\sum_{k=l}^{\infty}\binom{k+l-1}{l-1}g_{i,kn}$$
$$\leqslant\sum_{k=l}^{\infty}(k+l-1)^{l-1}g_{i,kn}, \tag{9}$$

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

$$\|u_{i,n}-\mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(m\bar{d_0})]\|_2=\left\|\sum_{k=l}^{\infty}\sum_{L_1+L_2+\cdots+L_l=k}(G_nA_nG_n)_{ii}^{k} -\mathrm{E}\left[\sum_{k=l}^{\infty}\sum_{L_1+L_2+\cdots+L_l=k}(G_nA_nG_n)_{ii}^{k}|\mathcal{F}_{i,n}(m\bar{d_0})\right]\right\|_2$$
$$\leqslant\sum_{k=l}^{\infty}(k+l-1)^{l-1}\|g_{i,kn}-\mathrm{E}[g_{i,kn}|\mathcal{F}_{i,n}(m\bar{d_0})]\|_2$$
$$\leqslant\sum_{k=l}^{\infty}(k+l-1)^{l-1}\sum_{j_1}\cdots\sum_{j_{k-1}}|a_{ij_1,n}\cdots a_{j_{k-1}i,n}| \cdot\|v_{i,n}^2\cdots v_{j_{k-1},n}^2-\mathrm{E}[v_{i,n}^2\cdots v_{j_{k-1},n}^2|\mathcal{F}_{i,n}(m\bar{d_0})]\|_2$$
$$\leqslant\sum_{k=l}^{\infty}(k+l-1)^{l-1}\eta^k<\int_{l-1}^{\infty}(x+l)^{l-1}\eta^x dx$$
$$\leqslant\int_{m-1}^{\infty}(x+l)^{l-1}\eta^x dx,$$

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,

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

Thus, for some constant $\overline{C_{1l}}>0$, $\int_{x=m-1}^{\infty}(x+l)^{l-1}\eta^x dx\leqslant\overline{C_{1l}}m^l\eta^m$. Therefore, we obtain $\|u_{i,n}-\mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(m\bar{d_0})]\|_2<\overline{C_{1l}}m^l\eta^m$.

When $m>l$,

$$\|u_{i,n}-\mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(m\bar{d_0})]\|_2=\left\|\sum_{k=l}^{\infty}\sum_{L_1+L_2+\cdots+L_l=k}(G_nA_nG_n)_{ii}^{k} -\mathrm{E}\left[\sum_{k=l}^{\infty}\sum_{L_1+L_2+\cdots+L_l=k}(G_nA_nG_n)_{ii}^{k}|\mathcal{F}_{i,n}(m\bar{d_0})\right]\right\|_2$$
$$\leqslant\sum_{k=l}^{m-1}(k+l-1)^{l-1}\|g_{i,kn}-\mathrm{E}[g_{i,kn}|\mathcal{F}_{i,n}(m\bar{d_0})]\|_2 +\sum_{k=m}^{\infty}(k+l-1)^{l-1}\|g_{i,kn}-\mathrm{E}[g_{i,kn}|\mathcal{F}_{i,n}(m\bar{d_0})]\|_2$$
$$\leqslant\sum_{k=l}^{m-1}(k+l-1)^{l-1}(2\eta C/(1-\eta))\eta^m+\sum_{k=m}^{\infty}(k+l-1)^{l-1}\eta^k$$
$$\leqslant(2\eta C/(1-\eta))\int_{2l-1}^{m+l-2}x^{l-1}dx\cdot\eta^m+\sum_{k=m}^{\infty}(k+l-1)^{l-1}\eta^k$$
$$<(2\eta C/(1-\eta))\cdot\frac{(m+l-2)^l-(2l-1)^l}{l}\eta^m +\int_{m-1}^{\infty}(x+l)^{l-1}\eta^x dx\leqslant\overline{C_{2l}}m^l\eta^m+\overline{C_{1l}}m^l\eta^m$$

for some constant $\overline{C_{2l}} > 0$. Thus $\|u_{i,n} - \mathrm{E}[u_{i,n}|\mathcal{F}_{i,n}(m\bar{d_0})]\|_2 \leqslant (\overline{C_{1l}} + \overline{C_{2l}})m^l\eta^m$. □

Lemma A.8. *Under Assumption 1, let $A_n = (a_{ij,n})$ be an $n \times n$ nonstochastic matrix with $|a_{ij,n}| \leqslant C_0 d(i,j)^{-\alpha}$ for some positive constants C_0 and α, where $d(i,j)$ is the distance between individuals i and j. Suppose $\sup_n \|A_n\|_\infty \leqslant \eta < 1$ and, for all positive integer numbers l, $\sup_n \||A_n|^l\|_1 \leqslant C_2 l\eta^l$ for some constant C_2, where $|A_n| \equiv (|a_{ij,n}|)_{n\times n}$. $\{v_{i,n}\}_{i=1}^n$ 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 $G_n = diag(v_{1,n},\ldots,v_{n,n})$. 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 $\overline{C_l} > 0$.

Proof of Lemma A.8. (i) Given any 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 $i \to j_1 \to j_2 \to \cdots \to j_{l-1} \to i$ 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 $\{a_{ij_1,n}, a_{j_1j_2,n}, \ldots a_{j_{l-1}i,n}\}$ that is $\leqslant C_0(s/l)^{-\alpha}$. Let $j_0 = i$. In $P(1)$, notice that

$$\begin{aligned}
&\|v_{i,n}^2v_{j_1,n}^2\cdots v_{j_{l-1},n}^2 - \mathrm{E}[v_{i,n}^2v_{j_1,n}^2\cdots v_{j_{l-1},n}^2|\mathcal{F}_{i,n}(s)]\|_2\\
&\leqslant \sum_{h=0}^{l-1}\|v_{j_h,n}^2 - \mathrm{E}[v_{j_h,n}^2|\mathcal{F}_{i,n}(s)]\|_2\\
&\leqslant \sum_{h=0}^{l-1}\|v_{j_h,n}^2 - \mathrm{E}^2[v_{j_h,n}|\mathcal{F}_{i,n}(s)]\|_2\\
&\leqslant 2\sum_{h=0}^{l-1}\|v_{j_h,n} - \mathrm{E}[v_{j_h,n}|\mathcal{F}_{j_h,n}(s-hs/l)]\|_2\\
&\leqslant 2C_1l^p\sum_{h=0}^{l-1}(l-h)^{-p}\cdot s^{-p}, \qquad (10)
\end{aligned}$$

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

$$\begin{aligned}
&\sum_{P(2)}|a_{ij_1,n}a_{j_1j_2,n}\cdots a_{j_{l-1}i,n}| \leqslant [(A_{1n}+A_{2n})^l]_{ii} - (A_{2n}^l)_{ii}\\
&\leqslant C_0(s/l)^{-\alpha}\sum_{h=0}^{l-1}\|A_{2n}\|_\infty^h\||A_n|^{l-h-1}\|_1\\
&\leqslant \left[C_0l^\alpha\sum_{h=0}^{l-1}\|A_n\|_\infty^hC_2(l-h-1)\eta^{l-h-1}\right]s^{-\alpha}\\
&\leqslant \left[C_0C_2\eta^{l-1}l^\alpha\sum_{h=0}^{l-1}(l-h-1)\right]s^{-\alpha}\\
&\leqslant \left[C_0C_2\eta^{l-1}l^{\alpha+1}(l-1)/2\right]s^{-\alpha} = C_{2l}s^{-\alpha}, \qquad (11)
\end{aligned}$$

where the first inequality follows from Lemma A.3. Hence, with Eq. (10),

$$\begin{aligned}
&\|g_{i,ln} - \mathrm{E}[g_{i,ln}|\mathcal{F}_{i,n}(s)]\|_2 \leqslant \sum_{P(2)}|a_{ij_1,n}a_{j_1j_2,n}\cdots a_{j_{l-1}i,n}|\\
&\quad + \sum_{P(1)}|a_{ij_1,n}a_{j_1j_2,n}\cdots a_{j_{l-1}i,n}|\\
&\quad \cdot\|v_{i,n}^2v_{j_1,n}^2\cdots v_{j_{l-1},n}^2 - \mathrm{E}[v_{i,n}^2v_{j_1,n}^2\cdots v_{j_{l-1},n}^2|\mathcal{F}_{i,n}(s)]\|_2\\
&\leqslant C_{2l}s^{-\alpha} + \eta^l\cdot 2C_1l^p\sum_{h=0}^{l-1}(l-h)^{-p}\cdot s^{-p} \leqslant C_ls^{-p}, \qquad (12)
\end{aligned}$$

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

(ii) By Eqs. (9), (11) and (12),

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

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

Lemma A.9. (1) *Under Assumption 2, if $\sup_{1\leqslant k\leqslant K,i,n}\mathrm{E}|x_{ik,n}|^p < \infty$ and $\sup_{i,n}\mathrm{E}|\epsilon_{i,n}|^p < \infty$ for some $p \geqslant 2$, 2, 1 and 2 respectively, then $\{z_{i,n}^2(\theta)\}_{i=1}^n$, $\{\ln\Phi(z_{i,n}(\theta))\}_{i=1}^n$, $\{\frac{\phi(z_{i,n}(\theta))}{\Phi(z_{i,n}(\theta))}\}_{i=1}^n$ and $\{\frac{\phi(z_{i,n}(\theta))z_{i,n}(\theta)}{\Phi(z_{i,n}(\theta))}\}_{i=1}^n$ are respectively uniformly (in i and n) $L_{p/2}$, $L_{p/2}$, L_p and $L_{p/2}$ bounded.*

(2) *Under Assumptions 1–3(1), if $\sup_{1\leqslant k\leqslant K,i,n}\mathrm{E}|x_{ik,n}|^p < \infty$ and $\sup_{i,n}\mathrm{E}|\epsilon_{i,n}|^p < \infty$ for some $p > 4$, 4, 6 and 8 respectively, then $\{z_{i,n}^2(\theta)\}_{i=1}^n$, $\{\ln\Phi(z_{i,n}(\theta))\}_{i=1}^n$, $\{\frac{\phi(z_{i,n}(\theta))}{\Phi(z_{i,n}(\theta))}\}_{i=1}^n$ and $\{\frac{\phi(z_{i,n}(\theta))z_{i,n}(\theta)}{\Phi(z_{i,n}(\theta))}\}_{i=1}^n$ are uniformly (in i and n) and geometrically L_2-NED on $\{x_{i,n},\epsilon_{i,n}\}_{i=1}^n$ with NED coefficients $\zeta^{s(p-4)/((2p-4)\bar{d_0})}$, $\zeta^{s(p-4)/((2p-4)\bar{d_0})}$, $\zeta^{s(p-6)/((2p-6)\bar{d_0})}$ and $\zeta^{s(p-8)/((2p-8)\bar{d_0})}$, respectively.*

(3) *Under Assumptions 1–3(2), if $\sup_{1\leqslant k\leqslant K,i,n}\mathrm{E}|x_{ik,n}|^p < \infty$ and $\sup_{i,n}\mathrm{E}|\epsilon_{i,n}|^p < \infty$ for some $p > 4$, 4, 6 and 8 respectively, then $\{z_{i,n}^2(\theta)\}_{i=1}^n$, $\{\ln\Phi(z_{i,n}(\theta))\}_{i=1}^n$, $\{\frac{\phi(z_{i,n}(\theta))}{\Phi(z_{i,n}(\theta))}\}_{i=1}^n$ and $\{\frac{\phi(z_{i,n}(\theta))z_{i,n}(\theta)}{\Phi(z_{i,n}(\theta))}\}_{i=1}^n$ are uniformly (in i and n) L_2-NED on $\{x_{i,n},\epsilon_{i,n}\}_{i=1}^n$ with NED coefficients $s^{-(\alpha-d)(p-4)/(2p-4)}$, $s^{-(\alpha-d)(p-4)/(2p-4)}$, $s^{-(\alpha-d)(p-6)/(2p-6)}$ and $s^{-(\alpha-d)(p-8)/(2p-8)}$, respectively.*

Proof of Lemma A.9. First, we consider $\{z_{i,n}^2(\theta)\}_{i=1}^n$. 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 Proposition 1 and Lemma A.4 with a there to be 1. Second, we consider $\{\ln\Phi(z_{i,n}(\theta))\}_{i=1}^n$. 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 $\lim_{x\to-\infty}g(x)/x = \lim_{x\to-\infty}\phi(x)/[\Phi(x)x] = \lim_{x\to-\infty}\frac{-x\phi(x)}{\Phi(x)+x\phi(x)} = \lim_{x\to-\infty}\frac{-1}{1+\Phi(x)/[x\phi(x)]} = -1$. Thus $|g(x)| < 2|x| + C_1$ for some constant $C_1 > 0$. By the mean value theorem, 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 Proposition 1, 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 Proposition 1 and Lemma A.4.

Third, we consider $\{\phi(z_{i,n}(\theta))/\Phi(z_{i,n}(\theta))\}_{i=1}^n$. Recall $|g(x)| < 2|x| + C_1$, thus $\{g(z_{i,n}(\theta))\}_{i=1}^n$ 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 Proposition 1 and Lemma A.4.

Finally, we study $\{\phi(z_{i,n}(\theta))z_{i,n}(\theta)/\Phi(z_{i,n}(\theta))\}_{i=1}^{n}$. Let $h(x) = \phi(x)x/\Phi(x)$. Because $g(x) < 2|x| + C_1$ for some constant $C_1 > 0$. Thus $|xg(x)| \leqslant 2x^2 + C_1|x|$ and by Proposition 1, $\{\frac{\phi(z_{i,n}(\theta))z_{i,n}(\theta)}{\Phi(z_{i,n}(\theta))}\}_{i=1}^{n}$ 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 Proposition 1 and Lemma A.4, we have the conclusion. □

Appendix B. Proofs

Proof of Lemma 1. When $\lambda_0 = 0$, the conclusion is trivial. Thus, in the remaining proof, we assume $\lambda_0 \neq 0$. In the following proof, for any matrix $A = (a_{ij})$, denote $|A| = (|a_{ij}|)$. 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 $\Gamma = |\lambda_0| \sup_n \|W_n\|_1 < \infty$ comes from

$$\sup_n \|W_n\|_1 = \sup_{n,j} \sum_{i=1}^{n} |w_{ij,n}| = \sup_{n,j} \sum_{m=1}^{\infty} \sum_{i:m\leqslant d(i,j)<m+1} |w_{ij,n}|$$
$$\leqslant \sup_j \sum_{m=1}^{\infty} \sum_{i:m\leqslant d(i,j)<m+1} C_0 m^{-\alpha} \leqslant \sum_{m=1}^{\infty} Cm^{d-1}C_0 m^{-\alpha} < \infty.$$

To simplify notations, we will use W_n to replace $|W_n|$ in the rest of the proof, because the analysis will work on upper bounds of inequalities. $\|W_n^l\|_1 = \max_{i=1,\ldots,n,} \|W_n^l e_i\|_1$, where $e_i = (0,\ldots,0,1,0,\ldots,0)'$ is the ith unit vector of dimension n. Let $\iota_n = \sum_{i=1}^{n} e_i$. Then $\|W_n^l e_i\|_1 = \iota_n'|W_n^l e_i|$. Without loss of generality, assume that only the first N columns satisfy $\Gamma \geqslant |\lambda_0| \sum_i |w_{ij,n}| > \zeta$ and the remaining $n-N$ columns satisfy $|\lambda_0| \sum_i |w_{ij,n}| \leqslant \zeta$. Then, for $l \geqslant 1$,

$$\begin{aligned}
\iota_n'|W_n^l e_i| &= (\iota_n' w_{\cdot 1,n}, \ldots, \iota_n' w_{\cdot n,n})|W_n^{l-1} e_i| \\
&= (\iota_n' w_{\cdot 1,n}, \ldots, \iota_n' w_{\cdot N,n}, 0, \ldots, 0)|W_n^{l-1} e_i| \\
&\quad + (0,\ldots,0, \iota_n' w_{\cdot (N+1),n}, \ldots, \iota_n' w_{\cdot n,n})|W_n^{l-1} e_i| \\
&\leqslant \|(\iota_n' w_{\cdot 1,n}, \ldots, \iota_n' w_{\cdot N,n}, 0, \ldots, 0)'\|_1 \|W_n^{l-1} e_i\|_\infty \\
&\quad + \|(0,\ldots,0, \iota_n' w_{\cdot (N+1),n}, \ldots, \iota_n' w_{\cdot n,n})'\|_\infty \|W_n^{l-1} e_i\|_1 \\
&\leqslant N|\lambda_0^{-1}|\Gamma \|W_n^{l-1} e_i\|_\infty + \lambda_m^{-1}\zeta \|W_n^{l-1} e_i\|_1 \\
&\leqslant N|\lambda_0^{-1}|\Gamma \|W_n^{l-1}\|_\infty \|e_i\|_\infty + \lambda_m^{-1}\zeta \|W_n^{l-1} e_i\|_1 \\
&\leqslant N|\lambda_0^{-l}|\Gamma \zeta^{l-1} + |\lambda_0^{-1}|\zeta \|W_n^{l-1} e_i\|_1.
\end{aligned}$$

Thus $\|W_n^l e_i\|_1 \leqslant N|\lambda_0^{-l}|\Gamma\zeta^{l-1} + |\lambda_0^{-1}|\zeta\|W_n^{l-1}e_i\|_1$. With recursion and $\|W_n e_i\|_1 \leqslant |\lambda_0^{-1}|\Gamma$, we obtain that $\|\lambda_0^l W_n^l e_i\|_1 \leqslant [(l-1)N + 1]\Gamma\zeta^{l-1} \leqslant \max(lN, 1)\Gamma\zeta^{l-1}$. □

Proof of Proposition 1. As before, for any matrix $A = (a_{ij})$, we denote $|A| = (|a_{ij}|)$.

(1): Denote the solution of $Y_n = F(\lambda_0 W_n Y_n + X_n\beta_0 + \epsilon_n)$ as $Y_n(X_n\beta_0 + \epsilon_n)$. Clearly, $Y_n(0) = 0$. Even though $F(x) = \max(0, x)$ is not differentiable, we can apply the mean value theorem of a convex function (Wegge, 1974) to $F(\cdot)$: $Y_n(X_n\beta_0 + \epsilon_n) - Y_n(0) = \overline{f_{D_n}}[\lambda_0 W_n(Y_n(X_n\beta_0 + \epsilon_n) - Y_n(0)) + X_n\beta_0 + \epsilon_n]$, where $\overline{f_{D_n}}$ is a diagonal matrix whose ith diagonal element is some subgradient of $F(\cdot)$ at some points between $\lambda w_{i\cdot,n} y_{i,n}(X_n\beta_0 + \epsilon_n)$ and 0. Thus $Y_n(X_n\beta_0 + \epsilon_n) - Y_n(0) = (I_n - \lambda_0\overline{f_{D_n}}W_n)^{-1}\overline{f_{D_n}}(X_n\beta_0 + \epsilon_n)$. Notice subgradients of $F(\cdot)$ are always between 0 and 1 and therefore, $(I_n - \lambda_0\overline{f_{D_n}}W_n)^{-1}\overline{f_{D_n}} = \sum_{l=0}^{\infty} \lambda_0^l(\overline{f_{D_n}}W_n)^l\overline{f_{D_n}} \leqslant^* \sum_{l=0}^{\infty} |\lambda_0 W_n|^l = (I_n - |\lambda_0 W_n|)^{-1} \equiv M_n \equiv (m_{ij,n})$, where $A \leqslant^* B$ means $|a_{ij}| \leqslant |b_{ij}|$ for all i's and j's. Hence, $|y_{i,n}(X_n\beta_0 + \epsilon_n)| \leqslant \sum_{j=1}^{n} m_{ij,n}|x_{j,n}\beta + \epsilon_{j,n}|$. Furthermore, by Minkowski's inequality, $\|y_{i,n}(X_n\beta_0 + \epsilon_n)\|_p \leqslant \sum_{j=1}^{n} m_{ij,n}\|x_{j,n}\beta + \epsilon_{j,n}\|_p$. With the uniform L_p-boundedness of $\{y_{i,n}\}_{i=1}^{n}$, the uniform L_p-boundedness of $\{w_{i\cdot,n}Y_n\}_{i=1}^{n}$, $\{z_{i,n}(\theta)\}_{i=1}^{n}$ and $\{y_{i,n}^*\}_{i=1}^{n}$ are a direct result of Minkowski's inequality.

(2) and (3): We first discuss the NED properties of $\{y_{i,n}\}_{i=1}^{n}$. Let $Y_n^{(1)} = F(\lambda_0 W_n Y_n^{(1)} + X_n^{(1)}\beta_0 + \epsilon_n^{(1)})$ and $Y_n^{(2)} = F(\lambda_0 W_n Y_n^{(2)} + X_n^{(2)}\beta_0 + \epsilon_n^{(2)})$. Then $Y_n^{(1)} - Y_n^{(2)} = \widetilde{f_{D_n}}[\lambda_0 W_n(Y_n^{(1)} - Y_n^{(2)}) + (X_n^{(1)} - X_n^{(2)})\beta_0 + (\epsilon_n^{(1)} - \epsilon_n^{(2)})]$, where $\widetilde{f_{D_n}}$ is a diagonal matrix whose ith diagonal element is some subgradient of $F(\cdot)$ at some point between $\lambda_0 w_{i\cdot,n} y_{i,n}^{(1)} + x_{j,n}^{(1)}\beta_0 + \epsilon_{j,n}^{(1)}$ and $\lambda_0 w_{i\cdot,n} y_{i,n}^{(2)} + x_{j,n}^{(2)}\beta_0 + \epsilon_{j,n}^{(2)}$. Thus, $Y_n^{(1)} - Y_n^{(2)} = (I_n - \lambda_0\widetilde{f_{D_n}}W_n)^{-1}\widetilde{f_{D_n}}[(X_n^{(1)} - X_n^{(2)})\beta_0 + (\epsilon_n^{(1)} - \epsilon_n^{(2)})]$. As in part (1), $(I_n - \lambda_0\widetilde{f_{D_n}}W_n)^{-1}\widetilde{f_{D_n}} \leqslant^* M_n$. Then by Prop. 1 in Jenish and Prucha (2012) and its proof, $\|y_{i,n} - \mathrm{E}[y_{i,n}|\mathcal{F}_{i,n}(s)]\|_2 \leqslant \sup_{j,n}\|\epsilon_{j,n} + x_{j,n}\beta_0\|_2 \sup_{i,n}\sum_{j:d(i,j)>s} m_{ij,n}$. We discuss the two cases in this proposition in the following:

(i) Under Assumption 3(1),

$$\begin{aligned}
\sup_{i,n} \sum_{j:d(i,j)>s} m_{ij,n} &= \sup_{i,n} \sum_{j:d(i,j)>s} \sum_{l} |\lambda_0|^l \cdot |(|W_n|^l)_{ij}| \\
&\leqslant \sup_{i,n} \sum_{j:d(i,j)>s} \sum_{l=[s/\bar{d}_0]+1} |\lambda_0|^l \cdot |(|W_n|^l)_{ij}| \\
&= \sup_{i,n} \sum_{l=[s/\bar{d}_0]+1} \sum_{j:d(i,j)>s} |\lambda_0|^l \cdot |(|W_n|^l)_{ij}| \\
&\leqslant \sup_{i,n} \sum_{l=[s/\bar{d}_0]+1} \zeta^l \leqslant (1-\zeta)^{-1}\zeta^{s/\bar{d}_0}.
\end{aligned}$$

(ii) When $\lambda_0 = 0$, it is trivial. Otherwise, under Assumption 3(2), 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.3,

$$\begin{aligned}
&\sum_{k_1}\cdots\sum_{k_{l-1}} |w_{ik_1,n}w_{k_1k_2,n}\cdots w_{k_{l-2}k_{l-1},n}w_{k_{l-1}j,n}| \\
&= (|W_n|^l)_{ij} = [(W_{Bn} + W_{An})^l]_{ij} - (W_{Bn}^l)_{ij} \\
&\leqslant C_0(d(i,j)/l)^{-\alpha} \sum_{h=0}^{l-1} \|W_{Bn}\|_\infty^h \|(|W_n|^{l-h-1})\|_1 \\
&\leqslant C_0(d(i,j)/l)^{-\alpha} \sum_{h=0}^{l-1} \|W_n\|_\infty^h C_1(l-h-1)\left(\frac{\zeta}{\lambda_0}\right)^{l-h-1}, \qquad (13)
\end{aligned}$$

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

$$\begin{aligned}
|(I_n - |\lambda_0 W_n|)_{ij}^{-1}| &= \sum_{l=1}^{\infty} |\lambda_0 W_n|_{ij}^l \\
&= \sum_{l=1}^{\infty} |\lambda_0^l| \sum_{k_1}\cdots\sum_{k_{l-1}} |w_{ik_1,n}w_{k_1k_2,n}\cdots w_{k_{l-1}j,n}| \\
&\leqslant C_0C_1 \sum_{l=1}^{\infty} |\lambda_0^l|(d(i,j)/l)^{-\alpha} \sum_{h=0}^{l-1} \|W_n\|_\infty^h (l-h-1)(\zeta/|\lambda_0|)^{l-h-1} \\
&\leqslant C_0C_1|\lambda_0| \sum_{l=1}^{\infty} (d(i,j)/l)^{-\alpha} \sum_{h=0}^{l-1} (l-h-1)\zeta^{l-1}
\end{aligned}$$

$$= C_0C_1|\lambda_0|\sum_{l=1}^{\infty}(d(i,j)/l)^{-\alpha}\zeta^{l-1}\frac{l(l-1)}{2}$$

$$\leqslant C_0C_1|\lambda_0|d(i,j)^{-\alpha}\sum_{l=1}^{\infty}l^{2+\alpha}\zeta^{l-1}/2 = C_2d(i,j)^{-\alpha}$$

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

$$\sup_{i,n}\sum_{j:d(i,j)>s}m_{ij,n} \leqslant \sup_{i,n}\sum_{m=[s]}^{\infty}\sum_{j:m\leqslant d(i,j)<m+1}C_2d(i,j)^{-\alpha}$$

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

$$\leqslant \sum_{m=[s]}^{\infty}C_3C_2(m+1)^{d-1}[(m+1)/2]^{-\alpha} \leqslant C_3C_22^{\alpha}\int_s^{\infty}x^{-\alpha+d-1}dx$$

$$= C_3C_22^{\alpha}(\alpha-d)^{-1}s^{d-\alpha},$$

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

Next, we discuss the NED of $\{w_{i\cdot,n}Y_n\}_{i=1}^n$ by the two settings of the weight matrix respectively under Assumption 3(1) and Assumption 3(2).

(i) Under Assumption 3(1), 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

$$\|w_{i\cdot,n}Y_n - \mathrm{E}[w_{i\cdot,n}Y_n|\mathcal{F}_{i,n}(m\bar{d}_0)]\|_2$$

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

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

$$\leqslant \sum_{j=1}^n|w_{ij,n}|\cdot\sup_{j,n}\|\epsilon_{j,n}+x_{j,n}\beta_0\|_2(1-\zeta)^{-1}\zeta^{m-1}$$

$$\leqslant \sup_{j,n}\|\epsilon_{j,n}+x_{j,n}\beta_0\|_2(1-\zeta)^{-1}\lambda_m^{-1}\zeta^m.$$

(ii) By Lemma A.1 in Jenish and Prucha (2009), $|\{j : m \leqslant d(i,j) < m+1\}| \leqslant C_1m^{d-1}$ for some constant $C_1 > 0$. When s is large enough,

$$\|w_{i\cdot,n}Y_n - \mathrm{E}[w_{i\cdot,n}Y_n|\mathcal{F}_{i,n}(s)]\|_2$$

$$\leqslant \sum_{k:d(k,i)\leqslant s/2}|w_{ik,n}|\cdot\|y_{k,n} - \mathrm{E}[y_{k,n}|\mathcal{F}_{i,n}(s)]\|_2$$

$$+ \sum_{k:d(k,i)>s/2}|w_{ik,n}|\cdot\|y_{k,n} - \mathrm{E}[y_{k,n}|\mathcal{F}_{i,n}(s)]\|_2$$

$$\leqslant \sum_{k:d(k,i)\leqslant s/2}|w_{ik,n}|\cdot\|y_{k,n} - \mathrm{E}[y_{k,n}|\mathcal{F}_{k,n}(s/2)]\|_2$$

$$+ \sum_{m=[s/2]}^{\infty}\sum_{k:m\leqslant d(k,i)<m+1}|w_{ik,n}|\cdot\|y_{k,n}\|_2$$

$$\leqslant \sum_{k:d(k,i)\leqslant s/2}|w_{ik,n}|C(s/2)^{d-\alpha} + (\sup_{n,k}\|y_{k,n}\|_2)\sum_{m=[s/2]}^{\infty}C_1m^{d-1}C_0m^{-\alpha}$$

$$\leqslant \|W_n\|_\infty C(s/2)^{d-\alpha} + (\sup_{n,k}\|y_{k,n}\|_2)C_1C_02^{\alpha-d+1}\int_{[s/2]}^{\infty}\frac{dx}{x^{\alpha-d+1}}$$

$$\leqslant C_2s^{d-\alpha}$$

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

Finally, the NED properties of $\{z_{i,n}(\theta)\}_{i=1}^n$ and $\{y_{i,n}^*\}_{i=1}^n$ are obvious from $\{w_{i\cdot,n}Y_n\}_{i=1}^n$. □

Proof of Proposition 2. For any $\epsilon > 0$, let $B = \{|y_{i,n}^*| < \epsilon, |\mathrm{E}[y_{i,n}^*|\mathcal{F}_{i,n}(s)]| < \epsilon\}$. Since $|\mathbb{I}(x_1 > 0) - \mathbb{I}(x_2 > 0)| \leqslant \frac{|x_1-x_2|}{\epsilon}\mathbb{I}(|x_1| > \epsilon \text{ or } |x_2| > \epsilon) + \mathbb{I}(|x_1| < \epsilon, |x_2| < \epsilon)$ (see the proof of Proposition 1 of Lei, 2013),

$$\|\mathbb{I}(y_{i,n}^* > 0) - \mathrm{E}[\mathbb{I}(y_{i,n}^* > 0)|\mathcal{F}_{i,n}(s)]\|_2$$

$$\leqslant \|\mathbb{I}(y_{i,n}^* > 0) - \mathbb{I}\{\mathrm{E}[y_{i,n}^*|\mathcal{F}_{i,n}(s)] > 0\}\|_2$$

$$\leqslant \left\{\int_{B^c}[y_{in}^* - \mathrm{E}(y_{in}^*|\mathcal{F}_{i,n}(s))]^2dP/\epsilon^2 + P(B)\right\}^{1/2}$$

$$\leqslant \frac{1}{\epsilon}\|y_{i,n}^* - \mathrm{E}[y_{i,n}^*|\mathcal{F}_{i,n}(s)]\|_2 + P(|y_{i,n}^*| < \epsilon)^{1/2}$$

$$\leqslant \frac{1}{\epsilon}\|y_{i,n}^* - \mathrm{E}[y_{i,n}^*|\mathcal{F}_{i,n}(s)]\|_2 + (C_2\epsilon)^{1/2},$$

for some constants $C_1 > 0$ and $C_2 > 0$, where the first inequality originates by Theorem 10.12 in Davidson (1994), the third inequality is based on the inequality $\sqrt{a+b} \leqslant \sqrt{a}+\sqrt{b}$ for arbitrary nonnegative numbers a and b, and 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 2. 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 $Y_{1,n}^* = \lambda_0W_{12,An}Y_{2,n} + X_1\beta_0 + \epsilon_{1,n}$ and $Y_{2,n}^* = Y_{2,n} = \lambda_0W_{22,An}Y_{2,n} + X_2\beta_0 + \epsilon_{2,n}$. 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 $(n-1)$ of the elements without i.

When $y_{i,n}^* > 0$, there are 2^{n-1} possible different A's with $i \in A \subset \{1, 2, \ldots, n\}$. Given each A, $Y_{2,n}^* = \lambda_0W_{22,An}Y_{2,n}^* + X_2\beta_0 + \epsilon_{2,n}$. Hence $Y_{2,n}^* = (I_{n_2} - \lambda_0W_{22,An})^{-1}(X_2\beta_0 + \epsilon_{2,n})$. That is to say, on such a regime A, $y_{i,n}^* \sim N((I_{n_2} - \lambda_0W_{22,An})_{i\cdot}^{-1}X_2\beta_0, \sigma_0((I_{n_2} - \lambda_0W_{22,An})^{-1}(I_{n_2} - \lambda_0W_{22,An}')^{-1})_{ii})$. By denoting the corresponding density function as $f_A(y_{i,n}^*)$, we have

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

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

Because the integral of a conditional density function is 1, $\sum_{i\in A\subset\{1,2,\ldots,n\}}\int f_A(Y_{-i,n}^*|y_{i,n}^*)dY_{-i,n}^* = 1$. 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 $\inf_{A,i,n}((I_{n_2} - \lambda_0W_{22,An})^{-1}(I_{n_2} - \lambda_0W_{22,An}')^{-1})_{ii} > 0$. 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 $x = (0, \ldots, 0, 1, 0, \ldots, 0)$ in the above inequality, where 1 locates in the jth position. We obtain $M_{jj} \geqslant \min eig(M), \forall j$. Hence,

$$\inf_{A,i,n}((I_{n_2} - \lambda_0W_{22,An})^{-1}(I_{n_2} - \lambda_0W_{22,An}')^{-1})_{ii}$$

$$\geqslant \inf_{A,i,n}\min eig((I_{n_2} - \lambda_0W_{22,An})^{-1}(I_{n_2} - \lambda_0W_{22,An}')^{-1})$$

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

where the second inequality comes from that the spectral radius of a matrix is less than or equal to its arbitrary norm, and the last inequality holds because $\sup_n \|W_n\|_1 < \infty$ under Assumption 3.

When $y^*_{i,n} < 0$, there are 2^{n-1} possible different A's where $A \subset \{1, 2, \ldots, n\} \setminus \{i\}$. 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 $N(x_{i,n}\beta_0, \sigma_0^2)$. When $A \neq \emptyset$, because $Y^*_{2,n} = (I_{n_2} - \lambda W_{22,An})^{-1}(X_2\beta_0 + \epsilon_{2,n})$,

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

Thus, given A, the relevant density for y^*_{it} 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 4. Let μ be a measure defined on $[0, \infty)$, $\mu([0, a]) = 1 + a$, and $\mu^n = \mu \otimes \cdots \otimes \mu$ be the product measure of n μ's. Because $\ln x \leqslant 2\sqrt{x} - 2$, $\mathrm{E}\ln[L_n(\theta)/L_n(\theta_0)] \leqslant 2\mathrm{E}(\sqrt{L_n(\theta)/L_n(\theta_0)} - 1) = 2\int(\sqrt{L_n(\theta)/L_n(\theta_0)} - 1)L_n(\theta_0)d\mu^n(Y_n) = 2(\int\sqrt{L_n(\theta)L_n(\theta_0)}d\mu^n(Y_n) - 1) = -\int[\sqrt{L_n(\theta)} - \sqrt{L_n(\theta_0)}]^2 d\mu^n(Y_n) \leqslant 0$. 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.

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

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

$$
\begin{aligned}
&\frac{n}{2}\ln(2\pi\sigma_0^2) - \ln|I_n - \lambda_0 W_n| \\
&\quad + \frac{1}{2}\sum_{i=1}^n \left(\frac{1}{\sigma_0} y_{i,n} - \frac{\lambda_0}{\sigma_0} w_{i\cdot,n} Y_n - x_{i,n}\frac{\beta_0}{\sigma_0}\right)^2 \\
&= \frac{n}{2}\ln(2\pi\sigma_1^2) - \ln|I_n - \lambda_1 W_n| \\
&\quad + \frac{1}{2}\sum_{i=1}^n \left(\frac{1}{\sigma_1} y_{i,n} - \frac{\lambda_1}{\sigma_1} w_{i\cdot,n} Y_n - x_{i,n}\frac{\beta_1}{\sigma_1}\right)^2 \qquad (14)
\end{aligned}
$$

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

$$
\begin{aligned}
&-\frac{\lambda_0}{\sigma_0^2}\sum_{i=1}^n (y_{i,n} - \lambda_0 w_{i\cdot,n} Y_n - x_{i,n}\beta_0) w_{ij,n} \\
&\quad + \frac{y_{j,n} - \lambda_0 w_{j\cdot,n} Y_n - x_{j,n}\beta_0}{\sigma_0^2} \\
&= -\frac{\lambda_1}{\sigma_1^2}\sum_{i=1}^n (y_{i,n} - \lambda_1 w_{i\cdot,n} Y_n - x_{i,n}\beta_1) w_{ij,n} \\
&\quad + \frac{y_{j,n} - \lambda_1 w_{j\cdot,n} Y_n - x_{j,n}\beta_1}{\sigma_1^2}. \qquad (15)
\end{aligned}
$$

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. (15) 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 $W_n + W_n' \neq 0$ and $w_{ii,n} \equiv 0$, we have $\lambda_0 = \lambda_1$. Eq. (15) 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 to show $\sup_{\theta\in\Theta}\frac{1}{n}|L_n(\theta) - \mathrm{E}L_n(\theta)| \xrightarrow{p} 0$ 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_1-NED random field is uniformly L_p bounded for some $p > 1$ on some suitable α-mixing base, then the weak law of large numbers (WLLN) holds. We have shown the uniform L_2-NED and uniform L_p boundedness of related terms of the Tobit model in the previous lemmas and propositions. By Eq. (3),

$$
\begin{aligned}
&\frac{1}{n}[\ln L_n(\theta) - \mathrm{E}\ln L_n(\theta)] = \frac{1}{n}\sum_{i=1}^n \{\mathbb{I}(y_{i,n} = 0)\ln\Phi(z_{i,n}(\theta)) \\
&\quad - \mathrm{E}[\mathbb{I}(y_{i,n} = 0)\ln\Phi(z_{i,n}(\theta))]\} \\
&\quad - \frac{1}{2n}\ln(2\pi\sigma^2)\sum_{i=1}^n [\mathbb{I}(y_{i,n} > 0) - \mathrm{E}\mathbb{I}(y_{i,n} > 0)] \\
&\quad + \frac{1}{n}(\ln|I_{2,n} - \lambda W_{22,n}| - \mathrm{E}\ln|I_{2,n} - \lambda W_{22,n}|) \\
&\quad - \frac{1}{2n}\sum_{i=1}^n \{\mathbb{I}(y_{i,n} > 0)z_{i,n}^2(\theta) - \mathrm{E}[\mathbb{I}(y_{i,n} > 0)z_{i,n}^2(\theta)]\}. \qquad (16)
\end{aligned}
$$

Because of the compactness of σ^2, the convergence to zero in probability of the second term on the right hand side of Eq. (16) 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 random fields, by Theorem 17.9 in Davidson (1994), $\{\mathbb{I}(y_{i,n} > 0)y_{i,n}^2\}_{i=1}^n$, $\{(w_{i\cdot,n}Y_n)^2\mathbb{I}(y_{i,n} > 0)\}_{i=1}^n$, $\{w_{i\cdot,n}Y_n\mathbb{I}(y_{i,n} > 0)\}_{i=1}^n$ and $\{(w_{i\cdot,n}Y_n)(x_{i,n}\beta)\}_{i=1}^n$ are all uniformly L_1-NED random fields. And by Lemma A.9, they are all uniformly $L_{2+\Delta/2}$ bounded, thus the pointwise WLLN is applicable to

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

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 $L_{1n}(\theta) \equiv \frac{1}{n}\sum_{i=1}^n\{\mathbb{I}(y_{i,n} = 0)\ln[1 - \Phi(\frac{\lambda}{\sigma}w_{i\cdot,n}Y_n + x_{i,n}\frac{\beta}{\sigma})] - \mathrm{E}[\mathbb{I}(y_{i,n} = 0)\ln[1 - \Phi(\frac{\lambda}{\sigma}w_{i\cdot,n}Y_n + x_{i,n}\frac{\beta}{\sigma})]]\}$ in Eq. (16). For any $\theta \in \Theta$, by Proposition 2, Lemmas 2 and A.9, $L_{1n}(\theta) \xrightarrow{p} 0$. By the compactness of the parameter space and Theorem 1 in Andrews (1992), it is sufficient to show that $L_{1n}(\theta)$ is SE. To do so, we only need to check the conditions of Corollary 3.1 in Andrews (1992). Let $\tilde{\lambda} = \frac{\lambda}{\sigma}$, $\tilde{\beta} = \frac{\beta}{\sigma}$, $\tilde{\sigma} = \sigma^{-1}$, similar to the reparameterization due to Olsen (1978). Because the parameter space is compact, $|\frac{\partial\tilde{\theta}_j}{\partial\theta_k}|$ is bounded for all j's and k's by simple calculation. 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.5, it suffices to show that $L_{1n}(\tilde{\lambda}, \tilde{\beta})$ is SE. Evidently, the ranges of $\tilde{\lambda}$ and $\tilde{\beta}$ are compact. Denote $\tilde{\lambda}_m = \sup\tilde{\lambda}$ and $\tilde{\beta}_m = \sup_{\tilde{\beta}}\max_{j=1}^K|\tilde{\beta}_j|$. Recall from the proof of Lemma A.9, $|\ln\Phi(x_1) - \ln\Phi(x_2)| \leqslant (2|x_1| + 2|x_2| + C_1)|x_1 - x_2|$ for some constant C_1. Hence,

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

By Proposition 1, $\sup_{i,n}\|w_{i\cdot,n}Y_n\|_{4+\Delta} < \infty$. Thus, $\sup_{i,n}\|\{4\tilde{\lambda}_m|w_{i\cdot,n}Y_n| + 4\tilde{\beta}_m\sum_{j=1}^K|x_{ij,n}| + C_1\}(|w_{i\cdot,n}Y_n| + \sum_{j=1}^K|x_{ij,n}|)\|_{2+\Delta/2} < \infty$ by Cauchy's inequality. Therefore, by Lemma 1 (a) in Andrews (1992), $L_{1n}(\theta)$ is SE 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 found in Qu and Lee (2013). Let $G_n(Y_n) = diag(\mathbb{I}(y_{1,n} > 0), \ldots, \mathbb{I}(y_{n,n} > 0))$. Then, as shown in Qu and Lee (2013),

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

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: $S_{i,n}(\lambda) \equiv \sum_{l=1}^{K_\epsilon}\frac{\lambda^l}{l}g_{il,n}$, where $g_{il,n} = \mathbb{I}(y_{i,n} > 0)\sum_{j_1}\cdots\sum_{j_{l-1}}w_{ij_1,n}w_{j_1j_2,n}\cdots w_{j_{l-2}j_{l-1},n}w_{j_{l-1}i,n}\mathbb{I}(y_{j_1,n} > 0)\cdots\mathbb{I}(y_{j_{l-1},n} > 0)$, and $R_{i,n}(\lambda) \equiv \mathbb{I}(y_{i,n} > 0)\sum_{l=K_\epsilon+1}^\infty\frac{\lambda^l}{l}g_{il,n}$. 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 A.8, 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 $\sup_{i,n}|g_{il,n}| < \infty$. Hence, $\frac{1}{n}\sum_{i=1}^n(g_{il,n} - \mathrm{E}g_{il,n}) \xrightarrow{p} 0$ follows from the WLLN in Jenish and Prucha (2012). Thus, $\sup_{\lambda\in\Lambda}|\frac{1}{n}\sum_{i=1}^n[S_{i,n}(\lambda) - \mathrm{E}S_{i,n}(\lambda)]| \xrightarrow{p} 0$.

Notice that there is a constant K_ϵ such that $\sup_{\lambda\in\Lambda}|\frac{1}{n}\sum_{i=1}^n R_{i,n}(\lambda)| \leqslant \sum_{l=K_\epsilon+1}^\infty\frac{\lambda_m^l}{l}\|W_n\|_\infty^l < \frac{\zeta^{K_\epsilon+2}}{K_\epsilon(1-\zeta)} < \frac{\epsilon}{4}$. Similarly, $\sup_{\lambda\in\Lambda}|\frac{1}{n}\sum_{i=1}^n\mathrm{E}R_{i,n}(\lambda)| < \frac{\epsilon}{4}$. Hence, $\sup_{\lambda\in\Lambda}|\frac{1}{n}\sum_{i=1}^n[R_{i,n}(\lambda) - \mathrm{E}R_{i,n}(\lambda)]| < \frac{\epsilon}{2}$. Consequently, as $n \to \infty$,

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

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

Previously, we have shown $\{\frac{1}{n}\sum_{i=1}^n\mathrm{E}[\mathbb{I}(y_{i,n} = 0)\ln\Phi(-\tilde{\lambda}w_{i\cdot,n}Y_n - x_{i,n}\tilde{\beta})]\}$ is equicontinuous. Because Θ is compact, 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.5, we only need to show that

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

is equicontinuous with respect to $\tilde{\theta}$. 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 $L_{2+\Delta/2}$ bounded, $\{x_{i,n}\}_{i=1}^n$ is uniformly $L_{4+\Delta}$ 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 $\mathrm{E}\ln|I_{2,n}-\lambda W_{22,n}| = \mathrm{E}\ln|I_n-\lambda\widetilde{W}_n|$. $\mathrm{E}\ln|I_{2,n}-\lambda W_{22,n}|/n$ is equicontinuous because

$$\sup_{\lambda,n}\left|\frac{d}{d\lambda}\frac{1}{n}\mathrm{E}\ln|I_{2,n}-\lambda W_{22,n}|\right| = \sup_{\lambda,n}\left|\frac{1}{n}tr\left[(I_n-\lambda\widetilde{W}_n)^{-1}\widetilde{W}_n\right]\right| \leqslant \sup_n \mathrm{E}\sum_{l=0}^{\infty}\lambda_m^l\|\widetilde{W}_n\|_\infty^{l+1} < \infty. \quad \square$$

Proof of Proposition 5. By Corollary 1 in Jenish and Prucha (2012), with Assumptions 10 and 11, to show the CLT, it is sufficient to check the uniform $L_{2+\tilde{\delta}}$ integrability, where $\tilde{\delta}$ is defined in Assumption 10, the uniform NED property of $\{\|q_{i,n}(\theta_0)\|\}_{i=1}^n$ in Eq. (8), where $\|\cdot\|$ is the Euclidean vector norm, and the decreasing rate of the NED coefficient.

We discuss the NED property separately by the two different settings in Assumption 3. Under Assumption 3(1), $\{z_{i,n}(\theta)^2\}_{i=1}^n$, $\{\ln\Phi(z_{i,n}(\theta))\}_{i=1}^n$, $\{\frac{\phi(z_{i,n}(\theta))}{\Phi(z_{i,n}(\theta))}\}$, $\{\frac{\phi(z_{i,n}(\theta))w_{i\cdot,n}Y_n}{\Phi(z_{i,n}(\theta))}\}$ and $\{r_{i,n}(\lambda_0)\}_{i=1}^n$ are uniformly and geometrically L_2-NED random fields and uniformly $L_{4+\delta/2}$ bounded. 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 uniformly and geometrically L_2-NED random fields. By Lemma B.4 in Xu and Lee (2015) for the Euclidean norm, $\{\|q_{i,n}(\theta_0)\|\}_{i=1}^n$ is a uniformly and geometrically L_2-NED random field. Then conditions (c) and (d) in Assumption 4 in Jenish and Prucha (2012) are satisfied.

Under Assumption 3(2), from Proposition 3(2), $\{r_{i,n}(\lambda_0)\}_{i=1}^n$ is a uniformly L_2-NED random field with coefficient $1/s^{(\alpha-d)/3}$. With Assumptions 5 and 10(i), by Lemma A.9, $\{z_{i,n}(\theta)^2\}_{i=1}^n$, $\{\ln\Phi(z_{i,n}(\theta))\}_{i=1}^n$, $\{\frac{\phi(z_{i,n}(\theta))}{\Phi(z_{i,n}(\theta))}\}$ and $\{\frac{\phi(z_{i,n}(\theta))w_{i\cdot,n}Y_n}{\Phi(z_{i,n}(\theta))}\}$ are all uniformly $L_{4+\delta/2}$ bounded, and uniformly L_2-NED random fields with NED coefficients $s^{-(\alpha-d)(4+\delta)/(12+2\delta)}$, $s^{-(\alpha-d)(4+\delta)/(12+2\delta)}$, $s^{-(\alpha-d)(2+\delta)/(10+2\delta)}$ and $s^{-(\alpha-d)\delta/(8+2\delta)}$. Because $\frac{4+\delta}{12+2\delta} > \frac{2+\delta}{10+2\delta} > \frac{\delta}{8+2\delta}$, the above four random fields are all with NED coefficient $s^{-(\alpha-d)\delta/(8+2\delta)}$. From Proposition 2, $\{\mathbb{I}(y_{i,n}=0)\}_{i=1}^n$ is uniformly NED with coefficient $1/s^{(\alpha-d)/3}$. Then by Lemma A.2, $\{\mathbb{I}(y_{i,n}=0)\phi(z_{i,n})w_{i\cdot,n}Y_n/\Phi(z_{i,n})\}_{i=1}^n$ is a uniformly NED random field with NED coefficient $\{\max[s^{-(\alpha-d)/3}, s^{-(\alpha-d)\delta/(8+2\delta)}]\}^{\delta/(8+2\delta)}$. 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 $\{\max[s^{-(\alpha-d)/3}, s^{-(\alpha-d)\delta/(8+2\delta)}]\}^{\delta/(8+2\delta)}$. That is to say, all the terms in the score are uniformly NED random fields with NED coefficient $\{\max[s^{-(\alpha-d)/3}, s^{-(\alpha-d)\delta/(8+2\delta)}]\}^{\delta/(8+2\delta)}$. Hence, by Lemma B.4 in Xu and Lee (2015), $\{\|q_{i,n}(\theta_0)\|\}_{i=1}^n$ is a uniformly L_2-NED random field with coefficient $\{\max[s^{-(\alpha-d)/3}, s^{-(\alpha-d)\delta/(8+2\delta)}]\}^{\delta/(8+2\delta)}$. Condition (c) in Assumption 4 in Jenish and Prucha (2012) requires $\sum_{s=1}^{\infty}s^{d-1}\{\max[s^{-(\alpha-d)/3}, s^{-(\alpha-d)\delta/(8+2\delta)}]\}^{\delta/(8+2\delta)} < \infty$, i.e., $\alpha > (7+24\delta^{-1})d$ and $\alpha > (5+32\delta^{-1}+64\delta^{-2})d$.

Next, it remains to check the uniform $L_{2+\tilde{\delta}}$ integrability of $\{\|q_{i,n}(\theta_0)\|\}_{i=1}^n$. It is sufficient to show that $\sup_{i,n}\mathrm{E}\|q_{i,n}(\theta)\|^{4+\delta/2} < \infty$ since $\tilde{\delta} < 2+\delta/2$ from Assumption 10 (see Exercise 5.4, p. 54, Shorack, 2000). From the C_r-inequality,[15]

$$\mathrm{E}\|q_{i,n}(\theta)\|^{4+\frac{\delta}{2}} \leqslant (2+K)^{1+\frac{\delta}{4}}\mathrm{E}\left\{\left|\frac{\partial\ln L_n(\theta)}{\partial\lambda}\right|^{4+\frac{\delta}{2}} + \sum_{k=1}^{K}\left|\frac{\partial\ln L_n(\theta)}{\partial\beta_k}\right|^{4+\frac{\delta}{2}} + \left|\frac{\partial\ln L_n(\theta)}{\partial\sigma}\right|^{4+\frac{\delta}{2}}\right\}. \quad (17)$$

Notice $|r_{i,n}(\lambda)| \leqslant \sum_{l=1}^{\infty}|\lambda|^l\sum_{j_1}\cdots\sum_{j_l}|w_{ij_1,n}w_{j_1j_2,n}\cdots w_{j_li,n}| \leqslant \sum_{l=1}^{\infty}|\lambda|^l\|W_n\|_\infty^{l+1} \leqslant \lambda_m^{-1}\frac{\zeta^2}{1-\zeta}$. With Lemma A.9, every term on the right hand side of Eq. (17) is uniformly $L_{4+\delta/2}$ integrable. Hence, all the conditions for the CLT of Jenish and Prucha (2012) are satisfied, and the asymptotic normality of the normalized score vector follows. $\square$

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 $\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,\Sigma_0^{-1})$. In the following, we will prove the above claim. Let $\psi(x) \equiv \frac{d[\phi(x)/\Phi(x)]}{dx} = \frac{-x\phi(x)}{\Phi(x)} - \frac{\phi^2(x)}{\Phi^2(x)}$. Then the second derivatives are

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

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

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

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

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

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

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

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

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, consider $\psi(x)$ and $\{\psi(z_{i,n})\}_{i=1}^n$, where $z_{i,n} = z_{i,n}(\theta_0)$. Note that $\psi'(x) = \frac{(x^2-1)\phi(x)}{\Phi(x)} + \frac{3x\phi^2(x)}{\Phi^2(x)} + \frac{2\phi^3(x)}{\Phi^3(x)}$. Since $\lim_{x\to-\infty}\phi(x)/[x\Phi(x)] = -1$, we have $|\psi(x)| = |-x\phi(x)/\Phi(x) - \phi^2(x)/\Phi^2(x)| \leqslant 3x^2 + C$ and $|\psi'(x)| \leqslant 7|x|^3 + C$ for some constant C. Thus, $\{\psi(z_{i,n})\}_{i=1}^n$ is also uniformly $L_{4+\delta/2}$ bounded and $|\psi(x_1) - \psi(x_2)| \leqslant (7|x_1|^3 + 7|x_2|^3 + C)|x_1 - x_2|$. Because $\{z_{i,n}\}_{i=1}^n$ is uniformly $L_{8+\delta}$ bounded, by Lemma A.4, $\{\psi(z_{i,n})\}_{i=1}^n$ is also a uniformly L_2-NED random field. Second, consider the product terms in the second derivatives. From Lemma A.9, $\{\frac{\phi(z_{i,n})}{\Phi(z_{i,n})}\}_{i=1}^n$ is uniformly $L_{8+\delta}$ bounded, and uniformly L_2-NED. By Cauchy's inequality, the terms $\{\mathbb{I}(y_{i,n}=0)\psi(z_{i,n})(w_{i\cdot,n}Y_n/\sigma)^2\}_{i=1}^n, \ldots, \{\mathbb{I}(y_{i,n}>0)z_{i,n}^2 w_{i\cdot,n}Y_n\}_{i=1}^n$ in the second derivatives of the log-likelihood function, are all uniformly $L_{2+\delta/4}$ bounded, and uniformly L_1-NED similarly to Theorem 17.9 in Davidson (1994). Thus the WLLN applies for these NED random fields.

With the above results, it remains to show that $\{tr[(I_{2,n} - \lambda_0 W_{22,n})^{-1}W_{22,n}]^2 - \mathrm{E}tr[(I_{2,n} - \lambda_0 W_{22,n})^{-1}W_{22,n}]^2\}/n \xrightarrow{p} 0$. Notice that $tr[(I_{2,n} - \lambda_0 W_{22,n})^{-1}W_{22,n}]^2 = tr[(I_n - \lambda_0\widetilde{W}_n)^{-1}\widetilde{W}_n]^2$, 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 3(2). Then the WLLN follows from the uniform boundedness of its elements:

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

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

It is sufficient to show that $\{\frac{1}{n}\frac{\partial^2 \ln L_n(\theta)}{\partial\theta\partial\theta'}\}_{i=1}^n$ is SE. This is because $\{v_T(\cdot) : T \geqslant 1\}$ is SE at τ_0, if and only if, for any sequence $\{\hat\tau_T : T \geqslant 1\}$ that satisfies $\rho(\hat\tau_T, \tau_0) \xrightarrow{p} 0$, where $\rho(\cdot,\cdot)$ is a metric, $v_T(\hat\tau_T) - v_T(\tau_0) \xrightarrow{p} 0$ (Andrews, 1994).

Most terms but one in the second derivatives are SE by Lemma A.6. 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 SE 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 $|\psi(x_1) - \psi(x_2)| \leqslant (7|x_1|^3 + 7|x_2|^3 + C)|x_1 - x_2|$, $\{z_{i,n}(\theta)\}_{i=1}^n$ is uniformly (in i and n) $L_{8+\delta}$ bounded, and $\{(w_{i\cdot,n}Y_n)^2\}_{i=1}^n$ are uniformly $L_{4+\delta/2}$ bounded, thus the conditions of Lemma A.6 are satisfied. The only term that Lemma A.6 is not applicable 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 SE because

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

Appendix C. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jeconom.2015.05.004.

References

Abadir, K.M., Magnus, J.R., 2005. Matrix Algebra. Cambridge University Press, Cambridge.

Amaral, P., Anselin, L., 2011. Finite Sample Properties of Moran's I Test for Spatial Autocorrelation in Probit and Tobit Models—Empirical Evidence. GeoDa Center for Geospatial Analysis and Computation.

Amemiya, T., 1973. Regression analysis when the dependent variable is truncated normal. Econometrica 41, 997–1016.

Amemiya, T., 1974. Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. Econometrica 42, 999–1012.

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge, MA.

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

Andrews, D.W., 1994. Asymptotics for semiparametric econometric models via stochastic equicontinuity. Econometrica 62, 43–72.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Boston.

Autant-Bernard, C., LeSage, J.P., 2011. Quantifying knowledge spillovers using spatial econometric models. J. Reg. Sci. 51, 471–496.

Bajari, P., Hong, H., Ryan, S.P., 2010. Identification and estimation of a discrete game of complete information. Econometrica 78, 1529–1568.

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

Calvó-Armengol, A., Patacchini, E., Zenou, Y., 2009. Peer effects and social networks in education. Rev. Econ. Stud. 76, 1239–1267.

Chernozhukov, V., Hong, H., Tamer, E., 2007. Estimation and confidence regions for parameter sets in econometric models1. Econometrica 75, 1243–1284.

Chudik, A., Pesaran, M.H., Tosetti, E., 2011. Weak and strong cross-section dependence and estimation of large panels. Econom. J. 14, C45–C90.

Ciliberto, F., Tamer, E., 2009. Market structure and multiple equilibria in airline markets. Econometrica 77, 1791–1828.

Davidson, J., 1994. Stochastic Limit Theory: An Introduction for Econometricians. Oxford University Press, Oxford.

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, 11th International Workshop Spatial Econometrics and Statistics.

Flores-Lagunes, A., Schnier, K.E., 2012. Estimation of sample selection models with spatial dependence. J. Appl. Econometrics 27, 173–204.

Gallant, A.R., White, H., 1988. A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. Basil Blackwell, New York.

Jenish, N., 2012. Nonparametric spatial regression under near-epoch dependence. J. Econometrics 167, 224–239.

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

Jenish, N., Prucha, I.R., 2012. On spatial processes and asymptotic inference under near-epoch dependence. J. Econometrics 170, 178–190.

Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. J. Real Estate Financ. Econ. 17, 99–121.

Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. Internat. Econom. Rev. 40, 509–533.

Koch, M., 2012. Three Essays in Spatial Econometrics. Vienna University of Economics and Business.

Lahiri, S.N., 1996. On inconsistency of estimators based on spatial data under infill asymptotics. Sankhyā 403–417.

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

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

Lei, J., 2013, Smoothed Spatial Maximum Score Estimation of Spatial Autoregressive Binary Choice Panel Models. Working paper, Tilburg University.

LeSage, J.P., 2000. Bayesian estimation of limited dependent variable spatial autoregressive models. Geogr. Anal. 32, 19–35.

LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics. CRC press, Boca Raton.

Marsh, T.L., Mittelhammer, R.C., 2004. Generalized maximum entropy estimation of a first order spatial autoregressive model. Adv. Econom. 18, 199–234.

McDonald, J.B., Newey, W.K., 1988. Partially adaptive estimation of regression models via the generalized t distribution. Econometric Theory 4, 428–457.

Nelson, F.D., 1977. Censored regression models with unobserved, stochastic censoring thresholds. J. Econometrics 6, 309–327.

Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. Handb. Econometrics 4, 2111–2245.

Olsen, R.J., 1978. Note on the uniqueness of the maximum likelihood estimator for the Tobit model. Econometrica 46, 1211–1215.

Pötscher, B.M., Prucha, I.R., 1997. Dynamic Nonlinear Econometric Models: Asymptotic Theory. Springer, New York.

Qu, X., Lee, L.-F., 2012. LM tests for spatial correlation in spatial models with limited dependent variables. Reg. Sci. Urban Econ. 42, 430–445.

Qu, X., Lee, L.-F., 2013. Locally most powerful tests for spatial interactions in the simultaneous SAR Tobit model. Reg. Sci. Urban Econ. 43, 307–321.

Qu, X., Lee, L.-F., 2015. Estimating a spatial autoregressive model with an endogenous spatial weight matrix. J. Econometrics 184 (2), 209–232.

Rothenberg, T.J., 1971. Identification in parametric models. Econometrica 39, 577–591.

Rupasingha, A., Goetz, S.J., Debertin, D.L., Pagoulatos, A., 2004. The environmental Kuznets curve for US counties: A spatial econometric analysis with extensions. Pap. Reg. Sci. 83 (2), 407–424.

Sheldon, R., 2002. A First Course in Probability. Pearson Education India, Upper Saddle River, NJ.

Shorack, G.R., 2000. Probability for Statisticians. Springer, New York.

Smith, T.E., 2009. Estimation bias in spatial models with strongly connected weight matrices. Geogr. Anal. 41, 307–332.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica 26, 24–36.

Wegge, L.L., 1974. Mean value theorem for convex functions. J. Math. Econom. 1, 207–208.

Xu, X., Lee, L.-F., 2015. A spatial autoregressive model with a nonlinear transformation of the dependent variable. J. Econometrics 186, 1–18.