Sieve Maximum Likelihood Estimation of the Spatial Autoregressive Tobit Model*

Xingbai Xu and Lung-fei Lee[†]

October 18, 2015

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

Abstract

This paper extends the ML estimation of a spatial autoregressive Tobit model under normal disturbances in Xu and Lee (2015, Journal of Econometrics) to distribution-free estimation. We examine the sieve MLE of the model, where the disturbances are i.i.d. with an unknown distribution. This model can be applied to spatial econometrics and social networks when data are censored. We show that related variables are weakly dependent, or more precisely, spatial near-epoch dependent (NED). An important contribution of this paper is that we develop some exponential inequalities for spatial NED random fields, which are also useful in other (e.g., semiparametric) studies when spatial correlation exists. With these inequalities, we establish the consistency of the estimator. Asymptotic distributions of structural parameters of the model are derived from a functional central limit theorem and projection.

Simulations show that the sieve MLE can improve the finite sample performance upon misspecified normal MLEs, in terms of reduction in the bias and standard deviation. As an empirical application, we examine the school district income surtax rates in Iowa. Our results show that the spatial spillover effects are significant, but they may be overestimated if disturbances are restricted to be normally distributed.

JEL: C14, C21, C24, C63

Keywords: spatial autoregressive model; Tobit model; sieve maximum likelihood estimation; near-epoch dependence; social network

---

*An earlier version of this paper was presented in the 4th Shanghai Econometric Workshop and a workshop at OSU. We thank the audiences for their comments. We would also like to acknowledge valuable discussions and suggestions received from Herman J. Bierens and Xiaohong Chen.

†Corresponding author. Tel: (614) 292-5508. E-mail address: lee.1777@osu.edu

1. Introduction and Literature Review

There has been growing interest and development in nonlinear spatial models. Jenish (2012) studies locally linear regression estimation of spatial near-epoch dependent (NED) processes. A spatial autoregressive (SAR) model with a nonlinear transformation of the dependent variable is considered in Xu and Lee (XL hereafter) (2015a) in order to capture dependent variables taking a limited range such as a share variable. Lei (2013) explores the smoothed maximum score estimation of spatial autoregressive binary choice panel models. Qu and Lee (QL hereafter) (2015) investigate the estimation of an SAR model with an endogenous spatial weights matrix, which explores NED features.

Usually, some types of laws of large numbers (LLN) and central limit theorems (CLT) are required when large sample properties of nonlinear estimators are examined. Most studies about LLN and CLT for spatial processes are studied in lattices in $\mathbb{Z}^d$. For instance, a CLT for stationary α-mixing random fields on $\mathbb{Z}^d$ is in Doukhan (1994) and functional CLT's for strictly stationary mixing processes on $\mathbb{Z}^d$ are in Dedecker (2001). However, as pointed out in Jenish and Prucha (JP hereafter) (2009), such settings are not useful for spatial econometrics, as spatial economic units are not located in a regular lattice pattern. They develop the weak LLN (WLLN) and CLT for mixing random fields on $\mathbb{R}^d$. Subsequently, LLN and CLT for NED random fields are established in JP (2012).

The importance of Tobit models in the microeconometric literature requires little explanation and earlier studies of Tobit models are summarized in Amemiya (1985). Recently, the interest in SAR Tobit models has been increasing, as a spatial agent's rational decision may be subject to non-negativity constraints. SAR Tobit models have various applications to different fields in economics. For example, it can be used to model origin-destination flows (LeSage and Pace, 2009) and study community-based health insurances (Donfouet, Jeanty and Malin, 2012). As pointed out in QL (2012), there are two types of SAR Tobit models: the simultaneous 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 one $y_{i,n} = \max(0, y^*_{i,n})$, where $y^*_{i,n} = \lambda_0 \sum_{j=1}^n w_{ij,n} y^*_{i,n} + x_{i,n}\beta_0 + \epsilon_{i,n}$. LeSage (2000) studies the Bayesian estimation of the latent SAR

Tobit model. Subsequent studies on the latent SAR Tobit model can be found in, e.g., Marsh and Mittelhammer (2004), and Amaral and Anselin (2013).

In this paper, we focus on the simultaneous SAR Tobit model and just call it the SART model for simplicity. The SART model is not only of interest in spatial econometrics, but also useful in social network analysis, because it is also a model of complete information games with linear-quadratic payoff functions subject to non-negativity constraints (Ballester, Calvó-Armengol and Zenou, 2006, Calvó-Armengol, Patacchini and Zenou, 2009, Allouch, 2012, XL, 2015b, Jackson and Zenou, 2014, and Bramoullé, Kranton, and D'Amours, 2014). There has been some empirical and theoretical work on the SART model. Rupasingha, Goetz, Debertin and Pagoulatos (2004) investigate the environmental Kuznets curve for US counties by this model. Autant-Bernard and LeSage (2011) consider the Bayesian estimation of the model and apply it to study knowledge spillovers. More empirical applications are discussed in XL (2015b).

In addition to empirical studies, there are also some theorectical researches on this model. QL (2012, 2013) propose tests for spatial correlation. XL (2015b) establish the consistency and asymptotic normality of the maximum likelihood estimation (MLE) of the SART model under normal disturbances. However, in empirical studies, we may not know the distribution of the error terms. The parametric MLE based on a misspecified normal distribution will be inconsistent. Finite sample biases due to misspecified distributions have been observed in the Monte Carlo study in XL (2015b).

In this paper, we aim to relax the normal distribution assumption in XL (2015b) and adopt sieves to approximate the true distribution of disturbances for estimation. Sieve estimation has been studied by Chen (2007) and Bierens (2014). They have established consistency and asymptotic normality for sieve estimators with an independent sample or a stationary time series. Our SART model is a nonlinear model with spatial dependence across observations generated by the model. The asymptotic analysis on properties of our sieve estimator extends both literatures of nonlinear time series and sieve estimation.

The structure of this paper is as follows. In Section 2, we introduce the SART model and study NED properties of related variables and statistics. In Section 3, we show consistency of the sieve

MLE. In Section 4, the asymptotic normality of structural parameter estimates is investigated, and we propose an estimator for its asymptotic variance. In Section 5, Monte Carlo simulations are designed to compare finite sample properties of the sieve MLE and parametric MLEs under misspecified normal distributions. We reexamine the school district income surtax rates issue explored in QL (2012) and XL (2015b) in Section 6. In Appendix A, we develop exponential inequalities for weakly dependent random fields, including uniformly bounded and unbounded NED random fields. Appendix B provides features on sieve approximation used for the unknown distribution of the model. The proofs for Sections 2 and 3 are summarized in Appendix C. In Appendix D, we list the first order and second order derivatives of the log-likelihood function and study properties of these derivatives. The proof for Section 4 is shown in Appendix E.[1]

2. The SART Model and NED Process

2.1. The Model

Individual spatial units in an economy are assumed to be located in a region $D_N \subset D \subset \mathbb{R}^d$, where the cardinality of D_N satisfies $|D_N| = N$, which is the sample size. We use $\vec{i}$ to represent individual i's location in $\mathbb{R}^d$. The distance between individuals i and j is $d(i,j) \equiv d(\vec{i}, \vec{j})$. For two subsets of spatial units of sizes u and v, define their distance $d(\{i_1, \cdots, i_u\}, \{j_1, \cdots, j_v\}) \equiv d(\{\vec{i}_1, \cdots, \vec{i}_u\}, \{\vec{j}_1, \cdots, \vec{j}_v\}) \equiv \min_{m,l}\{d(\vec{i}_m, \vec{j}_l) : 1 \leqslant m \leqslant u, 1 \leqslant l \leqslant v\}$.

Assumption 1. *The distance $d(\vec{i}, \vec{j})$ between any two different individuals i and j is larger than or equal to a specific positive constant, without loss of generality, say, 1.*

Under Assumption 1, there exists a constant $C_d > 0$ such that the number of points in a ball of radius r is less than or equal to $C_d(\lfloor r \rfloor + 1)^d$ (Lemma A.1, JP, 2009). The SART model is $y^*_{i,N} = \lambda_0 \sum_{j=1}^N w_{ij,N} y_{j,N} + x_{i,N}\beta_0 + \epsilon_{i,N} = \lambda_0 w_{i\cdot,N} Y_N + x_{i,N}\beta_0 + \epsilon_{i,N}$ and $y_{i,N} = \max(0, y^*_{i,N})$, where $w_{i\cdot,N}$ is the ith row of the spatial weights matrix W_N, $x_{i,N} \in \mathbb{R}^{K^0}$ is a vector of exogenous

[1] A supplement file to accompany the Appendices with detailed arguments is available online.

regressors and $y_{i,N}$ is an observed outcome of unit i. In matrix form,

$$Y_N = \max(0,\ \lambda_0 W_N Y_N + X_N \beta_0 + \epsilon_N), \tag{1}$$

where $\max(0,\ (a_1, \cdots, a_N)') \equiv (\max(0, a_1), \cdots, \max(0, a_N))'$. Assume that $\epsilon_{i,N}$'s are i.i.d, with a cumulative distribution function (CDF) $F_0(\cdot)$ and a probability density function (PDF) $f_0(\cdot)$, where the subscript 0 refers to the true value, but their functional forms are unknown. The X_N is an $N \times K^0$ matrix of exogenous variables. The spatial (network) matrix W_N is specified to have some basic properties as usual for a linear SAR model:

Assumption 2. *The W_N is a non-stochastic nonzero constant matrix with non-negative entries and its diagonal elements are all zero. The sequence $\{W_N\}$ is uniformly bounded in row and column sum norms.*

Our SART model has the feature of a simultaneous equation Tobit model in Amemiya (1974). The coherency of this model relies on the following condition:

Assumption 3. $\Lambda = [-\lambda_m,\ \lambda_m]$ *is the parameter space of λ for some finite positive bound λ_m, and $\zeta \equiv \lambda_m \sup_N ||W_N||_\infty < 1$.*

Assumption 3 is a typical assumption in an SAR model. Under this assumption for the SART model, Eq. (1) has a unique solution for Y_N via the contraction mapping theorem. This assumption is also used in QL (2013) and XL (2015b). Under Assumption 2, $\sup_N ||W_N||_\infty < \infty$ and $\sup_N ||W_N||_1 < \infty$. Assumption 3 has then imposed restriction on the range of the interaction parameter λ. For the case that W_N is row normalized such that $||W_N||_\infty = 1$, then λ_m can be taken as a real value slightly less than 1. Assumption 3 is a familiar assumption for a stable linear SAR model. When Assumption 3 fails, rather strong interactions might have taken place and the system might not be stable. In that case, there could be no, multiple or even infinite solutions for Eq. (1). More discussion of the failure of Assumption 3 can be found in XL (2015b).

By rearranging the (arbitrary) ordering of units, we decompose $Y_N = (Y_{1N}', Y_{2N}')'$, where all elements in Y_{1N} are zero while all elements in $Y_{2N} = Y_{2N}^*$ are strictly positive. Conformably,

W_N can be decomposed with $W_N = \begin{pmatrix} W_{11,N} & W_{12,N} \\ W_{21,N} & W_{22,N} \end{pmatrix}$. Then Eq. (1) is equivalent to $Y_{1N}^* = \lambda_0 W_{12,N} Y_{2N} + X_{1N}\beta_0 + \epsilon_{1N}$ and $Y_{2N} = \lambda_0 W_{22,N} Y_{2N} + X_{2N}\beta_0 + \epsilon_{2N}$. From QL (2013), the log-likelihood function for this model is

$$\begin{aligned} \ln L_N(\lambda, \beta, f|X_N) = &\sum_{i=1}^{N} 1(y_{i,N} = 0) \ln F(-\lambda w_{i\cdot,N} Y_N - x_{i,N}\beta) + \ln |I_{2,N} - \lambda W_{22,N}| \\ &+ \sum_{i=1}^{N} 1(y_{i,N} > 0) \ln f(y_{i,N} - \lambda w_{i\cdot,N} Y_N - x_{i,N}\beta), \end{aligned} \tag{2}$$

where the unknown density $f(\cdot)$ becomes an extra functional parameter. To consider identification of this model with a finite sample, we follow Rothenberg (1971). From Rothenberg (1971), $\theta^0 = (\lambda_0,\, \beta_0',\, F_0(\cdot))$ is identifiable iff there is no $\theta = (\lambda,\, \beta',\, F(\cdot)) \neq \theta^0$ such that $L_N(\theta) = L_N(\theta^0)$ a.s..

Assumption 4. (1) *$\epsilon_{i,N}$'s are i.i.d. double arrays and they are independent of X_N. Its PDF $f_0(\epsilon) > 0$ for all $\epsilon \in \mathbb{R}$ and it is differentiable.*

(2.1) *When $K^0 = 1$,* support$(x_{1,N},\, x_{2,N}, \cdots, x_{N,N}) = \mathbb{R}^N$ *and $\beta_0 \neq 0$.* (2.2) *When $K^0 > 1$, denote $x_{i,N} = (x_{i1,N},\, x_{i,\sim,N})$ and $\beta_{01} \neq 0$.* support$(x_{i,1,N}|x_{i,\sim,N}) = \mathbb{R}$ *a.s. and* rank[var$(x_{i,\sim,N})$] $= K^0 - 1$.

Lemma 1. *Under Assumptions 2-4, the model is identifiable.*

If $f_0(\cdot)$ is known to be symmetric about the origin or more generally with its median being zero, then the intercept of the model can also be identified.

Lemma 2. *Let $x_{i1,N} \equiv 1$. Under Assumptions 2 and 3, the model is identifiable if the following conditions hold:*

(1) *$\epsilon_{i,N}$'s are i.i.d. double arrays with median at 0 and they are independent of X_N. Its PDF $f_0(\epsilon) > 0$ for all $\epsilon \in \mathbb{R}$ and it is differentiable.*

(2.1) *When $K^0 = 2$,* support$(x_{12,N},\, x_{22,N}, \cdots, x_{N2,N}) = \mathbb{R}^N$ *and $\beta_{0,2} \neq 0$.* (2.2) *When $K^0 > 2$, denote $x_{i,N} = (1,\, x_{i2,N},\, x_{i,\sim,N})$ and $\beta_{0,2} \neq 0$.* support$(x_{i2,N}|x_{i,\sim,N}) = \mathbb{R}$ *a.s. and* rank[var$(x_{i,\sim,N})$] $= K^0 - 2$.

2.2. Spatial Near-Epoch Dependence

To study the large sample properties of our sieve estimator, we first establish some moment conditions and weak dependence properties for related variables from the SART model. We utilize the concept of spatial NED (JP, 2012). It is known that NED process in time series can accommodate autoregressive time series, but mixing processes might not. As an SAR model is an autoregressive process in space, it is natural to explore spatial NED properties. To do so, we need more assumptions.

Assumption 5. *The parameter space of β is $\Theta_\beta = \prod_{k=1}^{K^0}[-\beta_{km},\, \beta_{km}]$.*

Assumption 6. $\sup_{\beta\in\Theta_\beta,i,N} \mathrm{E}\exp(\gamma_x|x_{i,N}\beta|) < \infty$ *for some finite constant $\gamma_x > 0$.*

Assumption 7. $\sup_N \mathrm{E}\exp(\gamma_\epsilon|\epsilon_{i,N}|) < \infty$ *for some finite constant $\gamma_\epsilon > 0$.*

From Lemma 2 in XL (2015b), under Assumptions 3, 4 (1), 6 and 7, $\{y_{i,N}\}_{i=1}^N$ is uniformly L_p bounded: $\sup_{i,N} \mathrm{E}|y_{i,N}|^p < \infty$. If we are only interested in uniformly bounded moments property for $\{y_{i,N}\}$, weaker moment conditions than Assumptions 6 and 7, e.g. Assumption 4 in XL (2015b), are sufficient. However, Assumptions 6 and 7 are needed so that exponential inequalities (Theorem A.2) for $y_{i,N}$ and $z_{i,N}(\lambda,\beta) \equiv y_{i,N} - \lambda w_{i\cdot,N}Y_N - x_{i,N}\beta$ can hold. Exponential inequalities are essential tools for the consistency of a sieve extremum estimator (Wooldridge and White, 1991).

Lemma 3. *Under Assumptions 2 - 7, for any γ such that $0 < \gamma \leqslant (1 - \sup_N ||\lambda_0 W_N||_\infty)(1 + \zeta)^{-1}\min(\frac{1}{2}\gamma_x, \gamma_\epsilon)$, one has $\sup_{i,N} \mathrm{E}\exp(\gamma|y_{i,N}|) < \infty$ and $\sup_{\lambda,\beta,i,N} \mathrm{E}\exp(\gamma|z_{i,N}(\lambda,\beta)|) < \infty$.*

In order to establish NED properties of $y_{i,N}$, we need some additional structures for the weights matrix.

Assumption 8. *Only individuals whose distances are less than or equal to some specific constant $\bar{d}_0 > 1$ may directly affect each other. That is to say, the element $w_{ij,N}$ of the weights matrix W_N can be non-zero only if $d(\vec{i},\vec{j}) \leqslant \bar{d}_0$.*

With these assumptions, as shown in XL (2015b), $\{y_{i,N}\}_{i=1}^N$ is a uniform and geometrical L_2-NED random field (in short, UG L_2-NED) on $\{x_{i,N},\, \epsilon_{i,N}\}$. Here we summarize those elementary properties for easy reference.

Proposition 1. *Let* $\mathcal{F}_{i,N}(s) \equiv \sigma(\{x_{j,N}, \epsilon_{j,N} : d(\vec{i}, \vec{j}) \leqslant s\})$. *Under Assumptions 1-8,*

(1) $||y_{i,N} - \mathrm{E}[y_{i,N}|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_y \zeta^{s/\bar{d}_0}$;

(2) $\sup_{i,N} ||w_{i.,N} Y_N - \mathrm{E}[w_{i.,N} Y_N|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_{WY} \zeta^{s/\bar{d}_0}$;

(3) $\sup_{\lambda,\beta,i,N} ||z_{i,N}(\lambda, \beta) - \mathrm{E}[z_{i,N}(\lambda, \beta)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_Z \zeta^{s/\bar{d}_0}$.

There are functions which transform the above NED random fields in our model, which can also be NED random fields from the following lemma, established as Lemma A.4 in XL (2015b).

Lemma 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* $||u_{i,N} - \mathrm{E}[u_{i,N}|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_2\psi(s)$ *for all* i *and* N, *and* $\sup_{i,N} ||u_{i,N}||_{L^p} < \infty$ *for some* $p > 2a+2$. *Then* $||G(u_{i,N}) - \mathrm{E}[G(u_{i,N})|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C\psi(s)^{(p-2a-2)/(2p-2a-2)}$.

From Eq. (2), we need to deal with the indicator of noncensoring $1(y_{i,N} > 0)$. For that purpose, from Proposition 2 in XL (2015b), a sufficient condition is that the densities of $y^*_{i,N}$ are uniformly bounded in i and N. Under normal disturbances, the uniform boundedness of the PDF of $y^*_{i,N}$ is established in Lemma 2 in XL (2015b). In the context of this paper, even though we do not know the distribution of the disturbance, we can establish the uniform boundedness of the densities of $y^*_{i,N}$ by convolution and thus obtain the NED of $\{1(y_{i,N} > 0)\}_{i=1}^N$ under the following assumption:

Assumption 9. *At least one of the following conditions holds:* (1) $\lambda_0 \geqslant 0$; (2) *if* $\lambda_0 < 0$, $\sup_N ||\lambda_0 W_N||_\infty < 0.7548$; (3) W_N *is symmetric or row-normalized from a symmetric matrix; or* (4) W_N *is a lower triangular or upper triangular matrix.*

Conditions (1) and (2) in Assumption 9 reflect positive and negative spatial effects respectively, while condition (4) covers the case in time series with an initial period.

Lemma 5. *Under Assumptions 1 - 4, 7 and 9, the essential supremums of densities of* $y^*_{i,N}$ *are uniformly bounded in* i *and* N.

Corollary 1. *Under Assumptions 1 - 9,* $||1(y_{i,N} > 0) - \mathrm{E}[1(y_{i,N} > 0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_{1(y>0)} \zeta^{s/3\bar{d}_0}$ *for some constant* $C_{1(y>0)}$.

3. The Sieve MLE and Its Consistency

Because $\int[\sqrt{f_0(x)}]^2dx = 1$, $\sqrt{f_0(\cdot)} \in L^2(\mathbb{R})$, which is a Hilbert space, and accordingly, $\sqrt{f_0(\cdot)}$ can be approximated by Hermite polynomials, as suggested in Gallant and Nychka (1987). However, with $f(\cdot)$ being a density constructed by Hermite polynomials and $F(\cdot)$ the corresponding distribution (e.g., as constructed by Gallant and Nychka 1987), possible NED properties of $\ln F(z_{i,N}(\lambda, \beta))$ and $\ln f(z_{i,N}(\lambda, \beta))$ are hard to be established due to enormous fluctuation of the logarithm of Hermite polynomials. As a result, the use of Hermite polynomials is not desirable for our model. To overcome this complication, a device based on a monotonic transformation in Bierens (2014) is adopted.

From Assumption 4, $f_0(\cdot)$ is positive and continuous on $\mathbb{R}$. Given a prior chosen strictly increasing distribution function $G(\cdot)$ with derivative $g(x) \equiv G'(x) > 0$ for all $x \in \mathbb{R}$, $F_0(x) = H_0(G(x))$ where $H_0(\cdot)$ is a distribution function on $[0, 1]$. For example, $G(x)$ can be the Logistic distribution $1/(e^{-x}+1)$, or the standard normal distribution $\Phi(x)$. With a $G(\cdot)$, the problem of an unknown $F_0(\cdot)$ on $\mathbb{R}$ is changed into seeking $H_0(\cdot)$ over $[0, 1]$. In this way, we have more choices of basis functions, such as Legendre polynomials, Fourier series and cosine series, to approximate functions on $[0, 1]$. As $G(\cdot)$ is strictly increasing, $G^{-1}(\cdot)$ exists. Since $F(x) = H(G(x))$, $H(u) = F(G^{-1}(u))$. The corresponding density of $H(u)$ is $h(u) = f(G^{-1}(u))/g(G^{-1}(u))$. Denote $h_0(u) = f_0(G^{-1}(u))/g(G^{-1}(u))$. Because $f_0(x) > 0$ and $g(x) > 0$ for all $x \in \mathbb{R}$, $h_0(u) > 0$ for all $u \in (0, 1)$. When $H(u) = u$, $g(x) = f_0(x)$, which means the prior $g(x)$ is exactly $f_0(x)$. How to choose $G(x)$ is discussed in Bierens (2014). Given $G(\cdot)$, the log-likelihood function in Eq. (2) can be written as

$$\begin{aligned} \ln L_N(\theta) = &\sum_{i=1}^{N} 1(y_{i,N} = 0) \ln H(G(-\lambda w_{i\cdot,N} Y_N - x_{i,N}\beta)) + \ln |I_{2,N} - \lambda W_{22,N}| \\ &+ \sum_{i=1}^{N} 1(y_{i,N} > 0)[\ln g(y_{i,N} - \lambda w_{i\cdot,N} Y_N - x_{i,N}\beta) + \ln h(G(y_{i,N} - \lambda w_{i\cdot,N} Y_N - x_{i,N}\beta))]. \end{aligned} \tag{3}$$

As in Bierens (2014), we also adopt the cosine functions as the basis to approximate $h(u)$. But, as we can see from Bierens (2014), it is possible for the expectation of the approximating log-likelihood function to be $-\infty$ for some θ. Bierens (2014) assumes that data are i.i.d. and the set of such θ's does not contain an open ball. Such a possible negative infinity problem also occurs in Gallant and Nychka (1987). To overcome this problem, Gallant and Nychka (1987) add a strictly positive term in the density: $\epsilon_0\phi(\cdot)$, where $\epsilon_0 > 0$ is a very small number (e.g. 10^{-20}) such that it is negligible in computation, and $\phi(\cdot)$ is a given strictly positive density function, for example, a normal density. We adopt this idea from Gallant and Nychka (1987). Theoretically, this amounts to only consider densities satisfying

Assumption 10. *There is a constant* $0 < \epsilon_0 \ll 1$ *such that* $h(u) > \epsilon_0$ *for all* $u \in (0, 1)$*.*

Because $h(u)$ is continuous, by Theorems 3.1 and 3.2 in Bierens (2014), there is a unique series representation using the cosine basis $\{1, \sqrt{2}\cos k\pi u, k = 1,\ 2, \cdots\}$ as follows[2]:

$$h(u) = h(u|\delta) = (1-\epsilon_0)\frac{(1+\sum_{k=1}^{\infty}\delta_k\sqrt{2}\cos k\pi u)^2}{1+\sum_{k=1}^{\infty}\delta_k^2} + \epsilon_0, \tag{4}$$

for $u \in [0, 1]$, where $\delta = (\delta_1, \delta_2, ...)$ is a sequence of unknown coefficients, and

$$\begin{aligned} H(u) = H(u|\delta) = &\int_0^u h(v|\delta)dv \\ =&u + \frac{1-\epsilon_0}{1+\sum_{k=1}^{\infty}\delta_k^2}\{2\sqrt{2}\sum_{k=1}^{\infty}\delta_k\frac{\sin(k\pi u)}{k\pi} + \sum_{k=1}^{\infty}\delta_k^2\frac{\sin(2k\pi u)}{2k\pi} \\ &+ 2\sum_{k=2}^{\infty}\sum_{m=1}^{k-1}\delta_k\delta_m\frac{\sin((k+m)\pi u)}{(k+m)\pi} + 2\sum_{k=2}^{\infty}\sum_{m=1}^{k-1}\delta_k\delta_m\frac{\sin((k-m)\pi u)}{(k-m)\pi}\}. \end{aligned} \tag{5}$$

Properties of $h(u|\delta)$ and $H(u|\delta)$ are summarized in Appendix B. Denote $\widetilde{W_N} = I_N(Y)W_N I_N(Y)$, where $I_N(Y) \equiv \text{diag}(1(y_{1,N} > 0), \cdots, 1(y_{N,N} > 0))$. From XL (2015b), $\ln|I_{2,N} - \lambda W_{22,N}| = -\sum_{i=1}^N[\sum_{k=1}^{\infty}\lambda^k k^{-1}(\widetilde{W_N}^k)_{ii}]$. Thus, $\ln L_N(\theta) = \sum_{i=1}^N L_{i,N}(\theta)$, where

$$\begin{aligned} L_{i,N}(\theta) =&1(y_{i,N}=0)\ln H(G(z_{i,N}(\lambda,\beta))|\delta) - \sum_{k=1}^{\infty}\lambda^k k^{-1}(\widetilde{W_N}^k)_{ii} + \\ &1(y_{i,N}>0)[\ln g(z_{i,N}(\lambda,\beta)) + \ln h(G(z_{i,N}(\lambda,\beta))|\delta)]. \end{aligned} \tag{6}$$

[2]The theorems in Bierens (2014) apply directly to the density function $[h(u)-\epsilon_0]/(1-\epsilon_0)$.

Corollary 2.3 and Proposition 2.4 in White and Wooldridge (1991) provide sufficient conditions and steps for a general sieve estimator to be consistent. We shall verify those sufficient conditions for our sieve estimator of the SART model. We first define the parameter space of the structural parameters together with the sieve coefficients. For any $l \geqslant 0$, denote $||\delta||_l \equiv \sum_{i=1}^{\infty} i^l |\delta_i|$ and $||\theta||_l \equiv |\lambda| + \sum_{k=1}^{K^0} |\beta_k| + ||\delta||_l$ for any $\theta = (\lambda, \beta', \delta)$ as distances of parameter vectors from zero.

Assumption 11. *The true parameter vector* θ_0 *is in the parameter space* $\Theta \equiv \{(\lambda, \beta', \delta) : |\lambda| \leqslant \lambda_m, |\beta_k| \leqslant \beta_{km}, \forall k = 1, \cdots, K^0, ||\delta||_{l_0} \equiv \sum_{i=1}^{\infty} i^{l_0} |\delta_i| < \infty\}$ *for some* $l_0 \geqslant 1$.

$||\delta_0||_{l_0} < \infty$ imposes implicit restriction on h_0 as it controls sieve approximation in terms of series expansion via convergence behavior on the tail of series. The smallest l_0 to establish the consistency of our sieve estimator is 1, but for the asymptotic normality, $l_0 > d + 4$ is required. With sample size N, we choose $n - K^0$ (n depends on N) basis functions, $\{\cos k\pi u,\ k = 0, 1, \cdots, (n - 1 - K^0)\}$, to approximate the unknown $h_0(u)$. n is nondecreasing in N such that $\lim_{N\to\infty} n = \infty$, and the increasing rate will be determined later. Then, the parameter space of the model with a sample size N of the sieve estimation can be

$$\Theta_n = \{(\lambda, \beta', \delta) \in \Theta : ||\delta||_{l_0} \leqslant M_N, \delta_i = 0, \forall i > n - K^0\}, \tag{7}$$

where M_N satisfies $\lim_{N\to\infty} M_N = \infty$. Notice that Θ_n is a compact subset in $\mathbb{R}^n$ and $\cup_{n=1}^{\infty} \Theta_n$ is dense in Θ. The sieve MLE is $\hat{\theta}_n = \arg\max_{\theta \in \Theta_n} \ln L_N(\theta)$. Denote $Q_N(\theta) \equiv \mathrm{E} \ln L_N(\theta)$. From the information inequality and Lemma 1, $Q_N(\theta_0) > Q_N(\theta)$ for any $\theta \in \Theta$ but $\theta \neq \theta_0$. However, for consistency, we need to assume that the strictly inequality does not vanish as N tends to ∞:

Assumption 12. *For any* $\epsilon > 0$, $\liminf_{N\to\infty} \inf_{\theta \in \Theta : ||\theta - \theta_0||_{l_0} > \epsilon} \frac{1}{N} [Q_N(\theta_0) - Q_N(\theta)] > 0$.

In addition, in order to keep NED properties under transformation and apply the WLLN in JP (2012), we need additional structures on $g(\cdot)$. The exogenous variables are in general allowed to be spatially dependent.

Assumption 13. (1) $g(\cdot)$, $g'(\cdot)$ *and* $g''(\cdot)$ *are bounded by a constant* C_g; (2) $\frac{g(x)}{G(x)}$, $\frac{g'(x)}{g(x)}$, $\left|\frac{g''(x)}{g'(x)}\right|$ *and* $\left|\frac{g'''(x)}{g''(x)}\right|$ *are bounded by* $c(|x| + 1)$ *for all* $x \in \mathbb{R}$ *for some constant* c.

Assumption 14. *$\{x_{i,N}\}_{i=1}^N$ 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 $\sum_{r=1}^{\infty} r^{d-1}\hat{\alpha}(r) < \infty$.*

For all smooth enough $g(x)$ whose tails behavior is proportional to $\exp(-b|x|^a)$ for some $a \in [1, 2]$ and $b > 0$, including both logistic distributions and normal distributions, Assumption 13 holds. Assumption 14, which is also used in XL (2015b), is a requirement for the base field of NED in order that WLLN in JP (2012) may hold. With the above assumptions, we are ready to state the consistency of the sieve estimator of θ_0.

Theorem 1. *Under Assumptions 1 - 14, $||\hat{\theta}_n - \theta_0||_{l_0} = o_p(1)$, if $\lim_{N\to\infty} n = \infty$ and*

$$\lim_{N\to\infty} \frac{N}{M_N^4 (\ln M_N)^{2d+6} n^{2d+6}} = \infty. \tag{8}$$

We note that Eq. (8) is a sufficient condition which validates an exponential inequality for spatial NED so that uniform convergence of $\frac{1}{N}\ln L_N(\theta)$ on Θ_n can be achieved. Such a uniform convergence property is important for establishing consistency of a sieve estimator.

4. Asymptotic Normality

For asymptotic distribution of the sieve estimator, additional regularity conditions are needed.

Assumption 15. *For some $\tilde{\delta} > 0$, the α-mixing coefficient of $\{x_{i,N}\}_{i=1}^N$ in Assumption 14 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 16. *$l_0 > d+4$. $\theta_n^0 \equiv (\lambda_0, \beta_0', \delta_{10}, \cdots, \delta_{n-K^0-1,0}, 0, \cdots) \in \Theta_n^{\text{Int}}$, where Θ_n^{Int} is the interior of $\Theta_n = \{(\lambda, \beta', \delta) \in \Theta : ||\delta||_{l_0} \leqslant M_N, \delta_i = 0, \forall i \geqslant n - K^0\}$. And n is chosen such that $n^{-(l_0-1)}\sqrt{N} = o_p(1)$.*

For example, if we choose $n \propto (N^{1/(2d+6)}/\ln N)$ and $M_N \propto \ln N$, then Eq. (8) and Assumption 16 are satisfied. Denote $\nabla_k \equiv \nabla_{\theta_k} \equiv \frac{\partial}{\partial\theta_k}$ and $\nabla_{k,m} = \partial^2/\partial\theta_k\partial\theta_m$. For each $k = 1, \cdots, n$, by the mean value theorem,

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\nabla_k L_{i,N}(\hat{\theta}_n)+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\nabla_k L_{i,N}(\theta^0)-\nabla_k L_{i,N}(\theta^0_n)] \\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\nabla_k L_{i,N}(\theta^0)+\sum_{m=1}^{n}[\frac{1}{N}\sum_{i=1}^{N}\nabla_{k,m}L_{i,N}(\theta^0_n+\gamma_k(\hat{\theta}_n-\theta^0_n))]\sqrt{N}(\hat{\theta}_{n,m}-\theta_{0,m})
\end{aligned}\tag{9}$$

for some $\gamma_k \in [0,1]$. Following Bierens (2014), Eq. (9) can be converted into a single equation in random function form used in Eq. (13) below. Let

$$\widehat{V_n}(u)=\sum_{k=1}^{n}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}[\nabla_k L_{i,N}(\theta^0)-\nabla_k L_{i,N}(\theta^0_n)]\right]\eta_k(u),\tag{10}$$

$$\widehat{Z_n}(u)=\sum_{k=1}^{n}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\nabla_k L_{i,N}(\theta^0)\right]\eta_k(u),\tag{11}$$

$$\hat{b}_{m,n}(u)=-\sum_{k=1}^{n}\left[\frac{1}{N}\sum_{i=1}^{N}\nabla_{k,m}L_{i,N}(\theta^0_n+\gamma_k(\hat{\theta}_n-\theta^0_n))\right]\eta_k(u),\tag{12}$$

where $\eta_k(u)=2^{-k}\sqrt{2}\cos k\pi u$. From Remark A in Bierens (2014), the above three functions are both elements of the metric space of continuous functions $C[0,1]$ with the sup norm and the Hilbert space $L^2(0,1)=\overline{\text{span}(\{\sqrt{2}\cos m\pi u\}_{m=0}^{\infty})}$. Before moving forward, we need some moment and NED properties about the derivatives of the log-likelihood function:

Proposition 2. (1) *Under Assumptions 2 - 7, 11, 10 and 13,* $\{\nabla_\lambda L_{i,N}(\theta^0)\}$ *and* $\{\nabla_{\beta_j} L_{i,N}(\theta^0)\}$ *are uniformly (in i and N)* L_p *bounded for any* $p \geqslant 1$*, while* $\{\nabla_{\delta_k} L_{i,N}(\theta^0)\}$ *is uniformly bounded.*

(2) *Under Assumptions 1 - 10 and 13, (i) for any* $\gamma \in (0, \frac{1}{8})$*, there is a constant* $C>0$*, such that for all i and N,* $||\nabla_\lambda L_{i,N}(\theta^0)-\mathrm{E}[\nabla_\lambda L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2}\leqslant C\zeta^{\gamma s/\bar{d}_0}$ *and* $||\nabla_{\beta_k} L_{i,N}(\theta^0)-\mathrm{E}[\nabla_{\beta_k} L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2}\leqslant C\zeta^{\gamma s/\bar{d}_0}$*. (ii) There is a constant* $\overline{C}>0$ *such that* $||\nabla_{\delta_k} L_{i,N}(\theta^0)-\mathrm{E}[\nabla_{\delta_k} L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2}\leqslant \overline{C}k\zeta^{s/3\bar{d}_0}$ *for all i, N and k.*

Because $\nabla_k \ln L_N(\hat{\theta}_n)=0$, Eq. (9) implies $\sum_{m=1}^{n}\hat{b}_{m,n}(u)\sqrt{N}(\hat{\theta}_{n,m}-\theta_{0,m})=\widehat{Z_n}(u)-\widehat{V_n}(u)$, where $\widehat{V_n}(u)$ is neglectable:

Lemma 6. $\sup_{0\leqslant u\leqslant 1}|\widehat{V_n}(u)| = o_p(1)$.

By Lemma 6, $\sum_{m=1}^{n}\hat{b}_{m,n}(u)\sqrt{N}(\hat{\theta}_{n,m}-\theta_{0,m}) = \widehat{Z_n}(u)+o_p(1)$. We will show that $\widehat{Z_n}\Rightarrow Z$ for some random element Z, where $\Rightarrow$ means weak convergence in $C[0,1]$. Define $\widetilde{Z_N}(u) = \sum_{k=1}^{\infty}[N^{-1/2}\sum_{i=1}^{N}\nabla_k L_{i,N}(\theta^0)]\eta_k(u)$, which has the summation over k from 1 to ∞. It can be compared with $\widehat{Z_n}(u)$ that has the summation on k from 1 to n:

Lemma 7. $\sup_{0\leqslant u\leqslant 1}|\widehat{Z_n}(u)-\widetilde{Z_N}(u)| = o_p(1)$.

With Lemma 7, we transform the problem $\widehat{Z_n}(u)\Rightarrow Z$ into $\widetilde{Z_N}\Rightarrow Z$. As we will see in the proof of Lemma 8, $\sup_{0\leqslant u_1,u_2\leqslant 1}\limsup_{N\to\infty}\mathrm{cov}(\widetilde{Z_N}(u_1),\widetilde{Z_N}(u_2))<\infty$, but the following condition is needed to study the weak convergence of $\widetilde{Z_N}$:

Assumption 17. *For any u_1, $u_2\in[0,1]$, $\Gamma(u_1,u_2)\equiv\lim_{N\to\infty}\mathrm{cov}(\widetilde{Z_N}(u_1),\widetilde{Z_N}(u_2))$ exists.*

Lemma 8. *Under Assumptions 1-3, 6-9, 15 and 17, $\widetilde{Z_N}\Rightarrow Z$, where Z is a zero-mean Gaussian process on $[0,1]$ with covariance function $\Gamma(u_1,u_2)=\mathrm{E}[Z(u_1)Z(u_2)]$ and $\sup_{0\leqslant u_1,u_2\leqslant 1}|\Gamma(u_1,u_2)|<\infty$. Consequently, $\sum_{m=1}^{n}\hat{b}_{m,n}\sqrt{N}(\hat{\theta}_{n,m}-\theta_{0,m})\Rightarrow Z$.*

Therefore, we have

$$\begin{aligned}
&\sum_{m=1}^{n}\hat{b}_{m,n}(u)\sqrt{N}(\hat{\theta}_{n,m}-\theta_{0,m})\\
=&(\hat{b}_{1,n}(u),\cdots,\hat{b}_{K^0+1,n}(u))\sqrt{N}(\hat{\lambda}_n-\lambda_0,\hat{\beta}_n'-\beta_0')'+\sum_{m=1}^{n-K^0-1}\hat{b}_{K^0+1+m,n}(u)\sqrt{N}(\hat{\delta}_{n,m}-\delta_{0,m}) \qquad (13)\\
=&\widehat{Z_n}(u)-\widehat{V_n}(u)\Rightarrow Z(u).
\end{aligned}$$

The asymptotic distributions of $\hat{\lambda}_n$ and $\hat{\beta}_n$ can be recovered by an orthogonal projection as in Bierens (2014). Project each $\hat{b}_{m,n}(u)$, where $m=1,\cdots,K^0+1$, on the space spanned by $\hat{b}_{K^0+2,n}(u),\cdots,\hat{b}_{n,n}(u)$; and denote their residuals as $\hat{a}_{m,n}(u)$. Let $\hat{a}_n(u)=(\hat{a}_{1,n}(u),\cdots,\hat{a}_{K^0+1,n}(u))'$. Because $\int_0^1\hat{a}_n(u)(\hat{b}_{1,n}(u),\cdots,\hat{b}_{K^0+1,n}(u))du=\int_0^1\hat{a}_n(u)\hat{a}_n(u)'du$ and $\int_0^1\hat{a}_n(u)(\hat{b}_{K^0+2,n}(u),\cdots,\hat{b}_{n,n}(u))du=$

$0_{K^0+1,\, n-K^0-1}$, we have

$$\int_0^1 \hat{a}_n(u)\hat{a}_n(u)'du\sqrt{N}(\hat{\lambda}_n - \lambda_0, \hat{\beta}_n' - \beta_o')' = \int_0^1 \hat{a}_n(u)[\widehat{Z_n}(u) - \widehat{V_n}(u)]du. \tag{14}$$

If there exists a nonstochastic function $a(u)$ such that $\int_0^1 [\hat{a}_n(u) - a(u)][\hat{a}_n(u) - a(u)]'du \xrightarrow{p} 0$ and $0 < \int_0^1 a(u)a(u)'du < \infty$, then $\int_0^1 \hat{a}_n(u)\hat{a}_n(u)'du \xrightarrow{p} \int_0^1 a(u)a(u)'du$ and $\int_0^1 \hat{a}_n(u)[\widehat{Z_n}(u) - \widehat{V_n}(u)]du \Rightarrow \int_0^1 a(u)Z(u)du$. As a result,

$$\sqrt{N}(\hat{\lambda}_n - \lambda_0, \hat{\beta}_n' - \beta_o')' \Rightarrow \left[\int_0^1 a(u)a(u)'du\right]^{-1} \int_0^1 a(u)Z(u)du. \tag{15}$$

As $\hat{b}_{m,n}(u)$'s in Eq. (12) are defined in terms of the second order derivatives of the log likelihood function, in order to establish the above convergence, sample averages of the second order derivatives converge to some well-defined limits are needed as intermediate steps. For that purpose, we first establish their NED properties:

Proposition 3. *Under Assumptions 1 - 10, 13 and 16, we have the following UG L_2-NED properties:*

(1) *For any $\gamma \in (0, \frac{1}{8})$, there exists a constant C such that for all $1 \leqslant j \leqslant K^0$ and $1 \leqslant k \leqslant K^0$,*

(i) $||\nabla_{\lambda,\lambda} L_{i,N}(\theta^0) - \mathrm{E}[\nabla_{\lambda,\lambda} L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C\zeta^{\gamma s/\bar{d}_0}$,

(ii) $||\nabla_{\lambda,\beta_k} L_{i,N}(\theta^0) - \mathrm{E}[\nabla_{\lambda,\beta_k} L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C\zeta^{\gamma s/\bar{d}_0}$,

(iii) $||\nabla_{\beta_j,\beta_k} L_{i,N}(\theta^0) - \mathrm{E}[\nabla_{\beta_j,\beta_k} L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C\zeta^{\gamma s/\bar{d}_0}$, *for each $k = 1, \cdots, K^0$.*

(2) *For any $\gamma \in (0, \frac{1}{8})$, there exists a constant C such that for all $1 \leqslant j \leqslant K^0$ and $k \in \mathbb{N}$,*

(i) $||\nabla_{\lambda,\delta_k} L_{i,N}(\theta^0) - \mathrm{E}[\nabla_{\lambda,\delta_k} L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant Ck^2\zeta^{\gamma s/\bar{d}_0}$,

(ii) $||\nabla_{\beta_j,\delta_k} L_{i,N}(\theta^0) - \mathrm{E}[\nabla_{\beta_j,\delta_k} L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant Ck^2\zeta^{\gamma s/\bar{d}_0}$.

(3) *There exists a constant C such that $||\nabla_{\delta_j,\delta_k} L_{i,N}(\theta^0) - \mathrm{E}[\nabla_{\delta_j,\delta_k} L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C(j + k)\zeta^{s/3\bar{d}_0}$ for each pair $(j,k) \in \mathbb{N}^2$.*

With Lemmas D.1 and D.2, by Lebesgue's dominated convergence theorem, the order of the expectation and derivatives can be exchanged, i.e., $\mathrm{E}\,\nabla_j L_{i,N}(\theta) = \nabla_j \mathrm{E}\, L_{i,N}(\theta)$ and $\mathrm{E}\,\nabla_{j,k} L_{i,N}(\theta) = \nabla_{j,k} \mathrm{E}\, L_{i,N}(\theta)$ for all positive integers j and k. Because Θ is a separable metric space (Kreyszig, 1978, p.23), by Arzela-Ascoli Lemma (Royden and Fitzpatrick, 2010, p.207), there is a subsequence J_N of N, such that $\lim_{N\to\infty} \frac{1}{J_N} \mathrm{E} \ln L_{J_N}(\theta) = L_\infty(\theta)$ pointwisely. However, for our analysis, relatively strengthened assumptions are needed.

Assumption 18. *(i)* $\lim_{N\to\infty} \frac{1}{N} \mathrm{E} \ln L_N(\theta) = L_\infty(\theta)$*; (ii)* $\lim_{N\to\infty} \frac{1}{N} \nabla_j \mathrm{E} \ln L_N(\theta) = \nabla_j L_\infty(\theta)$ *and* $\lim_{N\to\infty} \frac{1}{N} \nabla_{j,k} \mathrm{E} \ln L_N(\theta) = \nabla_{j,k} L_\infty(\theta)$ *for all relevant positive integers j and k.*

Assumption 19. $\nabla_{j,k} L_\infty(\theta^0) \neq 0$ *for at least one pair (j, k) with $k \geqslant K^0 + 2$.*

For i.i.d. samples, Assumption 18 holds trivially. It indicates not only the existence of limit functions of $\frac{1}{N} \mathrm{E} \ln L_N(\theta)$ and its first order and second order derivatives, but also that the limit functions are twice continuously differentiable (Young's theorem). Additionally, it confirms the exchangeability of limit and differentiation. Assumption 19 corresponds to Assumptions 6.6 (c) and 7.2 in Bierens (2014). It will be used to verify condition (c) in Lemma E.2 in order to establish the next Lemma 9.

In terms of expectations, denote $b_{m,N}(u) \equiv -\sum_{k=1}^\infty \mathrm{E}[\frac{1}{N} \nabla_{k,m} \ln L_N(\theta^0)] \eta_k(u)$ and $b_m(u) \equiv -\sum_{k=1}^\infty \nabla_{k,m} L_\infty(\theta^0) \eta_k(u)$. Project $b(\cdot) = (b_1(\cdot), \cdots, b_{K^0+1}(\cdot))'$ onto $\mathrm{span}(\{b_m(\cdot)\}_{m=K^0+2}^\infty)$ and denote its residual $a(\cdot) \in \mathbb{R}^{K^0+1}$.

Lemma 9. *Under Assumptions 1-14, 16, 18 and 19,* $\mathrm{plim}_{N\to\infty} \int_0^1 [\hat{a}_n(u) - a(u)][\hat{a}_n(u) - a(u)]' du = 0$.

Finally, the following assumption is needed to establish the asymptotic distribution of $\hat{\lambda}_n$ and $\hat{\beta}_n$. More original conditions (Assumptions 6.7 and 6.8, Bierens, 2014) can be presumed such that Assumption 20 holds. But those conditions are also very hard, if not impossible to establish, therefore we impose Assumption 20 directly.

Assumption 20. $0 < \int_0^1 a(u) a(u)' du < \infty$.

Summarizing the above discussion, we reach the asymptotic distribution for the structural parameter estimates.

Theorem 2. *Under Assumptions 1-20,*

$$\sqrt{N}(\hat{\lambda}_n - \lambda_0,\, \hat{\beta}_n' - \beta_o')' \Rightarrow \left[\int_0^1 a(u)a(u)'du\right]^{-1} \int_0^1 a(u)Z(u)du \sim N_{K^0+1}(0,\Sigma), \tag{16}$$

where $\Sigma = [\int_0^1 a(u)a(u)'du]^{-1}[\int_0^1 \int_0^1 a(u_1)\Gamma(u_1,\, u_2)a(u_2)'du][\int_0^1 a(u)a(u)'du]^{-1}$ *with* $\Gamma(u_1,u_2) \equiv \lim_{N\to\infty} \operatorname{cov}(\widetilde{Z_N}(u_1), \widetilde{Z_N}(u_2)) = \mathrm{E}[Z(u_1)Z(u_2)]$.

For inference, the asymptotic variance needs to be estimated. With $\hat{\theta}_n$, modify Eq. (12) to $\bar{b}_{m,n}(u) = -\sum_{k=1}^{n}[\frac{1}{N}\sum_{i=1}^{N}\nabla_{k,m}L_{i,N}(\hat{\theta}_n)]\eta_k(u)$ and let $\bar{a}_n(u) = (\bar{a}_{1,n}(u), \cdots, \bar{a}_{1+K^0,n}(u))$ be the residual of projecting $(\bar{b}_{1,n}(u), \cdots, \bar{b}_{1+K^0,n}(u))$ on the rest $\bar{b}_{m,n}(u)$'s. Clearly, Lemma 9 still holds; thus, $\int_0^1 \bar{a}_n(u)\bar{a}_n(u)'du \xrightarrow{p} \int_0^1 a(u)a(u)'du$. It remains to estimate $\int_0^1\int_0^1 a(u_1)\Gamma(u_1,\, u_2)a(u_2)'du_1du_2$. Let $\hat{\Gamma}_n(u_1,\, u_2) = -\frac{1}{N}\sum_{k=1}^{n}\sum_{m=1}^{n}\nabla_{k,m}\ln L_N(\hat{\theta}_n)\eta_k(u_1)\eta_m(u_2)$. An estimate for the asymptotic variance is

$$\hat{\Sigma}_n = \left[\int_0^1 \bar{a}_n(u)\bar{a}_n(u)'du\right]^{-1}\left[\int_0^1\int_0^1 \bar{a}_n(u_1)\hat{\Gamma}_n(u_1,\, u_2)\bar{a}_n(u_2)'du\right]\left[\int_0^1 \bar{a}_n(u)\bar{a}_n(u)'du\right]^{-1}. \tag{17}$$

Proposition 4. *Under Assumptions 1-14, 16, 18- 20,* (1) $\sup_{u_1,u_2\in[0,1]}|\hat{\Gamma}_n(u_1,\, u_2) - \Gamma(u_1,\, u_2)| = o_p(1)$*;* (2) $\hat{\Sigma}_n - \Sigma = o_p(1)$*.*

One may question whether the variance of the limit distribution should depend on $\eta_k(u)$ or not. As $\bar{a}_n(u)$ depends on $\eta_k(u)$, at a first sight, it seems that $\hat{\Sigma}_n$ would depend on $\eta_k(u)$. The following proposition shows that $\hat{\Sigma}_n$ does not really depend on $(\eta_1(u), \cdots, \eta_n(u))$, so long as they are orthogonal. Let $(\chi_1(\cdot),\, \chi_2(\cdot), \cdots)$ be an orthonormal basis for $L^2(0,1)$, $(\omega_1,\, \omega_2, \cdots)$ be a sequence of real numbers, and $\Lambda_n = \operatorname{diag}(\omega_1, \cdots, \omega_n)$ such that $\eta_k(u) = \omega_k\chi_k(u)$. Also let $\eta(u) = (\eta_1(u), \cdots, \eta_n(u))'$. Then $\int_0^1 \eta(u)\eta(u)'du = \Lambda_n^2$. $\bar{b}_{m,n}(u)$, $\bar{a}_n(u)$ and $\hat{\Sigma}_n$ may be calculated with the specific $2^{-k}\sqrt{2}\cos(k\pi u)$, they are not necessarily. We find that $\hat{\Sigma}_n$ does not depend on the choice of orthonormal basis and such a feature corresponds to Theorem 6.2 in Bierens (2014). Denote the Hessian matrix $\hat{H}_n = (\hat{H}_{km,n}) \equiv (\frac{1}{N}\nabla_{k,m}\ln L_N(\hat{\theta}_n))$.

Proposition 5. *Not matter what a sequence of orthogonal functions* $(\eta_1(u), \eta_2(u), \cdots)$ *is,* $\hat{\Sigma}_n = -(\hat{H}_n^{-1})_{(1:K^0+1),(1:K^0+1)}$*, where* $(\hat{H}_n^{-1})_{(1:K^0+1),(1:K^0+1)}$ *is the upper-left* $(K^0+1)\times(K^0+1)$ *block* *of* $\hat{H}_n^{-1}$.

5. Monte Carlo Simulation

In this section, we investigate finite sample properties of the proposed sieve estimator. The model in the simulation is $y_{i,N} = \max(0,\, \lambda_0 w_{i\cdot,N} Y_N + \beta_{10} + \beta_{20} x_{i,N} + \epsilon_{i,N})$. For better comparing the MC results to those of the parametric MLE in XL (2015b), we adopt same data generating processes. The true parameters are: $\lambda_0 = 0.5$, $\beta_{10} = -1$, $\beta_{20} = 2$ with $x_{i,N}$ being *i.i.d.* $\sim N(0.2, 0.25)$. We generate $\epsilon_{i,N}$ from two different distributions, a mixed normal distribution (half probability $N(8/\sqrt{17}, 4/17)$, half probability $N(-8/\sqrt{17}, 4/17)$), which has two peaks, and also a Laplace distribution with standard deviation 2. Spatial weights matrices, W_N's, are generated as follows. The connection relationship of the 3142 counties in the U.S. can be found in U.S. Dept. of Commerce, Bureau of the Census (1992). Thus, we have a 3142×3142 matrix W_0 whose elements are one if the corresponding counties are contiguous; otherwise, zero. When a sample size $N = 1000$, we generate a uniform random natural number i between 1 and 2143, then we use the entries of W_0 between the i^{th} and the $(i + N - 1)^{th}$ rows and between the i^{th} and the $(i + N - 1)^{th}$ columns to form an $N \times N$ submatrix. We row-normalize that submatrix to obtain W_N. Similarly, W_N's are obtained for sample sizes 200 and 500, except that W_0 now is generated from 760 counties in the 10 Upper Great Plains States[3], rather than from all the counties in the U.S. To do so, we can have more nonzero elements when a sample size is smaller. With data of W_N, X_N and ϵ_N, and designed values of parameters, we generate the data of Y_N by contraction mapping. The iteration stops to obtain a "fixed point" Y_N when $||Y_N - F(\lambda W_N Y_N + \beta_{10} + \beta_{20} X_N + \epsilon_N)||_\infty < 10^{-6}$.

We first estimate the model by the parametric ML, as if disturbances are normally distributed. Next, we do sieve MLEs with k cosine basis functions, where k can be a value of $2, 3, \cdots$ or 10.[4] The prior chosen $g(\cdot)$ is a logistic density with mean $\hat{\beta}_1$ and standard deviation $\hat{\sigma}$, where $\hat{\beta}_1$ and $\hat{\sigma}$

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

[4] When we let ϵ_0 be 10^{-10} or even smaller, we can hardly see any difference in the estimators. Thus, in the experiments, we just let it be zero.

are the parametric MLEs based on normal distribution. When $k = 2$, the starting points for λ and β_2 in optimization are also the parametric MLEs, and the starting points for δ_1 and δ_2 are zero; when $3 \leqslant k \leqslant 10$, the starting point is the estimate with $k-1$ basis functions and zero for the new variable δ_k. After these with each k taking value from 2 to 10, we choose k such that the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) is minimized. In addition to the above MCs, we also conduct some other designs. We report results under a fixed number of sieves , which corresponds to either 5 or 10 basis functions to examine their finite sample properties. We try to use the standard logistic distribution $g(x) = 1/(e^x + e^{-x} + 2)$ without adapting $\hat{\beta}_1$ and $\hat{\sigma}$ to investigate whether locations and scales of the $G(\cdot)$ transformation would matter or not. For the case where disturbances are symmetrically distributed, we know by Lemma 2, the intercept term is identifiable and the coefficients $\delta_{0k} = 0$ with k being odd numbers of a sieve approximation. Therefore, when a symmetric distribution of the DGP is assumed, we use the same basis functions to approximate the true density but impose $\delta_k = 0$ for odd k's in estimation. The number of sieves is also determined by AIC or BIC. Results are reported in columns under "Symmetry" of Tables 1 and 2.

We are interested in the finite sample performance of $\hat{\lambda}$, $\hat{\beta}_1$ and $\hat{\beta}_2$. The results under the mixed normal distributed disturbance for DGP are reported in Table 1. We see that biases of the parametric MLEs are rather large, but the sieve estimates, no matter whether chosen by AIC or BIC, have much smaller biases and root mean square errors (RMSE) than those of the parametric MLE. The differences for the estimates of β_2 are especially obvious. When the sample size is 1000, the bias of the parametric MLE $\hat{\beta}_2$ is about 103 times as large as that of the sieve estimator from AIC and 318 times as large as that from BIC, and the RMSE of the parametric MLE $\hat{\beta}_2$ is about eight times as large as those from the sieve estimation chosen by AIC and BIC. When the sample size is not large, AIC performs better than BIC; but as the sample size increases, the bias from BIC decreases quickly. When $N = 1000$, although AIC still performs better than BIC from the criterion of RMSE, the difference is very small. For the design with 10 sieve terms, its performance is almost the same as those chosen from AIC with varying numbers of sieve terms. On the other hand, while the estimates with 5 sieve terms have worse finite sample properties than those chosen by AIC or

BIC, but it is still better than the parametric MLE. With the standard logistics transformation as an alternative one, the estimators also perform well. Although they have larger biases than those in location-scale adapted ones, the difference decreases fast as sample size increases. Also, their RMSEs are very close to those of location-scale adjusted ones. Thus, it is not essential to adjust the location and scale, especially when the sample size is large. With symmetry, performances of estimates are similar to those without exploring the symmetry, but they can have much more precisely estimated intercepts than those of the parametric MLE.

The Laplace distribution is more similar in shape to the normal distribution than the mixed normal distribution and thus the biases are much less. But we can still see that the sieve estimation performs better, and such a difference in performance is clearer when sample size is larger. When $N = 1000$, RMSEs of the parametric MLEs $\hat{\lambda}$ and $\hat{\beta}_2$ are respectively 14% and 100% greater than those of the sieve estimates, no matter selected by AIC or BIC. And the performances of AIC and BIC are very similar, although BIC is slightly better. The estimates with 5 or 10 sieves are worse than those chosen by AIC or BIC. The performance with standard logistics transformation is very similar to the location-scale adjusted case. When symmetry in distribution of disturbance is known and imposed, its performances are also similar to those without imposing symmetry.

From these experiments, we conclude that (1) the number of sieves can be decided by AIC or BIC instead of a fixed number, and it seems that AIC is a better choice since its performance is more robust, especially under small samples; (2) sometimes we can obtain smaller biases if we adjust the location and the scale, but it is generally unnecessary; (3) we do not need to add symmetry on the disturbance for estimation if intercept is not of special interest, even if the disturbances are really symmetrically distributed. By imposing symmetry in sieve approximation, the performance of estimates does not significantly improved.

Next, we summarize the number of cosine functions used in the estimation in Table 3. As we can see, the mixed normal distribution requires more basis functions than the Laplace distribution, because the Laplace distribution is closer to the prior chosen logistic transformation. We note that from the statistics literature, AIC tends to overfit a model while BIC tends to underfit a model, because the penalty for adding one more variable in BIC is $\log(N)$ while that in AIC is only 2. Our experiments also reflex this point: AIC tends to choose more basis functions. We see that when disturbances are mixed-normally distributed, in more than 10% of the experiments, AIC uses 10 sieve functions. It is entirely possible that with more choices in sieve terms, we might have better estimates. Thus, we also let the maximum number of sieves be 15 instead of 10 in order to see any difference. The results from Table 4 indicate that differences are small: $\text{mean}(|\hat{\lambda}_{10} - \hat{\lambda}_{15}|)/\lambda_0 < 2\%$, $\sqrt{\text{mean}(\hat{\lambda}_{10} - \hat{\lambda}_{15})^2}/\lambda_0 < 5\%$, $\text{mean}(|\hat{\beta}_{10} - \hat{\beta}_{15}|)/\beta_{20} \leqslant 0.6\%$, and $\sqrt{\text{mean}(\hat{\beta}_{10} - \hat{\beta}_{15})^2}/\beta_{20} < 1.8\%$.

Finally, finite sample properties of the estimators of the asymptotic variance of the sieve MLE are studied in Table 5. Since the experiments are repeated 1000 times, we can calculate its empirical standard deviation (std), denoted "empirical" in Table 5, which could be regarded as an accurate approximation to the true std of estimates in a finite sample. We also calculate the asymptotic std in each repetition by Eq. (17) or Proposition 5, and display their mean under columns "theo" in Table 5. From Table 5, it seems that the theoretical std's underestimate the true ones, although not much. Under the mixed normal disturbances, the biases for std estimates for $\hat{\beta}$ decrease as the sample size increases, however, the result for $\hat{\lambda}$ is not satisfactory, has a downward bias between 20%-30%. Under Laplace disturbances, when we use BIC such biases are not large.

6. Empirical Studies

QL (2012, 2013) and XL (2015b) study tax policy competition among local governments in Iowa using the data of school district income surtax rates. In Iowa, this type of surtax ranges from 0% to 20%. In 2009, surtax rates in 18.3% of the total 361 school districts in Iowa were 0%. Thus, the SART model is suitable to study this problem with spatial autocorrelation. Some theoretical background, detailed descriptive statistics of data, and the data source can be found in QL (2012). QL (2012, 2013) test the existence of spatial autocorrelation in the SART model under the assumption that disturbances are i.i.d. normally distributed. XL (2015b) give asymptotic theory for the MLE and derive standard errors of coefficient estimates under the assumption of i.i.d. normal distribution. However, if the disturbances terms are not normally distributed, the test statistics or estimates in all those earlier papers might not be robust. So, there is a need to look into estimates that can be distributional free.

There are two different settings on spatial weights matrices. The first one, labelled "adjacency", has $w^*_{ij,n} = 1$ when school districts i and j share a border, otherwise $w^*_{ij,n} = 0$. The second setting, denoted as "county", has $w^*_{ij,n} = 1$ if school districts i and j are in the same county; otherwise $w^*_{ij,n} = 0$. In both settings, W_n is the matrix row-normalized from W^*_n.

Estimation results are summarized in Tables 6 and 7. From Table 6, the coefficients of the variable "college graduates" are both small and insignificant. Accordingly, we also run a regression without it and summarize the results in Table 7. We see that both AICs and BICs in Table 7 are less than those in Table 6. Thus, it is suitable to drop "college graduates". From both tables, AICs and BICs for parametric MLE are about 40% and 37% higher than those from the sieve estimation. Thus, the distribution of disturbances in this empirical example may be quite different from the normal one. With the parametric MLE, we reject $\lambda_0 = 0$ at the 1% level in the "adjacency" setting and 10% level in the "county" setting. But in the "adjacency" setting, both tables show that the parametric MLE overestimates the spatial effects, although we still reject $\lambda_0 = 0$ at the 5% level in sieve estimation. In the "county" setting, the sieve estimate $\hat{\lambda}$ is insignificant in Table 6, and although it is significant in Table 7, its value is smaller ($\hat{\lambda} = 0.0726$) than those in the "adjacency" setting (AIC: 0.1744, BIC: 0.1334). Thus, in the "county" setting, we can say either there is no spatial correlation, or it is small. Hence, we have a different conclusion by dropping the normal distribution assumption.

Summarizing the above analysis, we think the "adjacency" setting without the variables "college graduates" is the best model. From the simulation, when samples are small, and the disturbance is far away from the normal distribution, AIC has better small sample properties. Thus, we focus on the results by AIC. First, standard errors of the estimates are apparently smaller in sieve estimation chosen by AIC than those of the parametric MLE. That is to say, we have more precise estimates. Second, the signs of the coefficients of "white", "pupil/taxpayer" and "property rate" are the same, but parametric MLE overestimates their absolute values. Third, "over 65" is not significant in the parametric MLE or the sieve estimation selected by BIC, but it is significantly negatively in the sieve estimation selected by AIC.

7. Conclusion

This paper relaxes the assumption that error terms are normally distributed in the SART model in XL (2015b) and we consider distribution free estimation. We consider the sieve ML estimation for this model, and study asymptotic properties of the sieve estimator. To show the uniform convergence in probability of the sample average sieve log likelihood function, this paper has first developed some exponential inequalities for weakly dependent random fields, including NED random fields, on $\mathbb{R}^d$. With these exponential inequalities for NED random fields, consistency of the sieve MLE for the SART model is established. And we obtain asymptotic normality of the structural parameter estimates by a functional central limit theorem and a projection method.

This paper has studied the sieve MLE of the SART model. Although we obtain some asymptotic results, there are still several possible extensions that can be studied in the future. (1): Decreasing rates of the exponential inequalities developed in this paper are slower than that in Bernstein's inequality. If we can obtain a faster decreasing rate, we may obtain asymptotic properties for sieve MLE under weaker assumptions. (2): We assume that the true density, after transformation, is bounded away from zero, and this condition is repeatedly used in the proof. Although similar assumptions are used in the literature but ignored in application, it limits the theoretical generality of this model. How to relax this assumption is a possible future research topic. (3): It remains for us to obtain the asymptotic distribution of the sieve distribution estimate.

Appendices

A. Exponential Inequalities for Weakly Dependent Random Fields

In spatial econometrics, usually there are both spatial correlation and heteroscedasticity and spatial units are not located in a regular lattice. The spatial process is not stationary. To establish consistency of a sieve estimator via Theorem 2.5 in White and Wooldridge (1991), a key step is to show uniform convergence in probability. To do that, we need to establish some large deviation inequalities. Saulis and Statulevicius (1991) and White and Wooldridge (1991) have summarized and proved large deviation inequalities for independent and mixing stochastic processes. There is a literature on large deviation inequalities for random fields on $\mathbb{Z}^d$, see, e.g., Ko (2013). But in our setting and most empirical applications in spatial econometrics, individuals are located or living in $\mathbb{R}^d$, rather than $\mathbb{Z}^d$. To the best of our knowledge, Delyon (2009) is the unique paper that includes large deviation inequalities for mixing random fields on $\mathbb{R}^d$. It examines large deviation inequalities for mixing random field, but its conditions are not applicable to our SART model. Because in our model, dependent variables and related transformations are NED random fields, we need large deviation inequalities for NED random fields on $\mathbb{R}^d$. In this paper, we establish a general result on exponential inequality that includes both mixing and NED random fields.

Our discussion will be based on Assumption 1. Let $\{X_{i,N} : \vec{i} \in D_N\}_{N=1}^{\infty}$ be a weakly dependent random field satisfying the following regularity conditions.

Assumption A.1. *$\{X_{i,N} : \vec{i} \in D_N\}_{N=1}^{\infty}$ is uniformly bounded in i and N, i.e., $\sup_{i,N} ||X_{i,N}||_{\infty} \leqslant M$, which is a finite constant, and centered, i.e., $\mathrm{E}\, X_{i,N} = 0$ for all i and N.*

Assumption A.2. *There exist positive constants C_{cv} and a_{cv}, such that for all $\{\vec{i}_1, \cdots, \vec{i}_u\} \subseteq D_N$ and $\{\vec{j}_1, \cdots, \vec{j}_v\} \subseteq D_N$, $|\mathrm{cov}(X_{i_1,N} \cdots X_{i_u,N}, X_{j_1,N} \cdots X_{j_v,N})| \leqslant C_{cv} M^q e^q \exp(-a_{cv} r)$, where $r = d(\{\vec{i}_1, \cdots, \vec{i}_u\}, \{\vec{j}_1, \cdots, \vec{j}_v\}) \equiv \min_{m,l}\{d(\vec{i}_m, \vec{j}_l) : 1 \leqslant m \leqslant u, 1 \leqslant l \leqslant v\}$ and $q \equiv u + v$.*

Denote $S_N \equiv \sum_{i=1}^{N} X_{i,N}$. For any integer $q \geqslant 2$, let $A_q(N) \equiv \sum_{1 \leqslant i_1 \leqslant \cdots i_q \leqslant N} |\mathrm{E}\, X_{i_1,N} \cdots X_{i_q,N}|$. To bound $|\mathrm{E}\, S_N^q|$, it is sufficient to bound $A_q(N)$, because $|\mathrm{E}\, S_N^q| \leqslant q! A_q(N)$. The following lemma provides combinatorial techniques to partition $\{i_1, i_2, \cdots, i_q\}$ in order to evaluate $|\mathrm{E}\, X_{i_1,N} \cdots X_{i_q,N}|$.

Lemma A.1. *For q points $\{i_1, i_2, \cdots, i_q\}$ in $\mathbb{R}^d$, which are not completely overlapped, there exists at least one partition of the q points, satisfying $I_1 \cup I_2 = \{i_1, i_2, \cdots, i_q\}$, $I_1 \cap I_2 = \emptyset$, $I_1 \neq \emptyset$ and $I_2 \neq 0$, such that there exists a real number r such that $d(I_1, I_2) = r$ and $\cup_{i \in I_k} \overline{B(i, r/2)}$ is path-connected for both $k = 1$ and 2, where $\overline{B(i, r/2)} \equiv \{x \in \mathbb{R}^d : d(i, x) \leqslant r/2\}$.*

Proof of Lemma A.1: It suffices to consider the case where there are no overlapped points. Consider the balls $\overline{B(i_k, R)}$. When $R = a \equiv \max_{1 \leqslant k, j \leqslant q} d(i_k, i_j)$, for any point i_k, $\overline{B(i_k, R)} \supseteq \{i_1, i_2, \cdots, i_q\}$, thus $\cup_{i \in I_k} \overline{B(i_k, R)}$ is path-connected. When $R = b \equiv \min_{k \neq j} d(i_k, i_j)$, it is clear that for any $j \neq k$, $\overline{B(i_k, R/3)}$ and $\overline{B(i_j, R/3)}$ are not connected. Thus, when R decreases from a to b, there must be a critical value $r/2$, such that $\cup_{i \in \{i_1, i_2, \cdots, i_q\}} \overline{B(i, r/2)}$ is path-connected but, for any $R < r/2$, $\cup_{i \in \{i_1, i_2, \cdots, i_q\}} \overline{B(i, R)}$ is not path-connected. Now let $\epsilon > 0$ be a sufficiently small number. Since $\cup_{i \in \{i_1, i_2, \cdots, i_q\}} \overline{B(i, r/2 - \epsilon)}$ is not connected, it is a union of several connect sets: $\cup_{i \in I'_1} \overline{B(i, r/2 - \epsilon)}$, $\cup_{i \in I'_2} \overline{B(i, r/2 - \epsilon)}$, $\cdots$, and $\cup_{i \in I'_J} \overline{B(i, r/2 - \epsilon)}$, where $\cup_{j=1}^{J} I'_j = \{i_1, i_2, \cdots, i_q\}$ and $I'_j \cap I'_k = \emptyset$ for any $j \neq k$. With the critical case $R = r/2$, there are two balls $\overline{B(i_j, r/2)}$ and $\overline{B(i_k, r/2)}$ are tangent, where, without loss of generality, $i_j \in I_1 = I'_1$ and $i_k \in I_2 = \cup_{j=2}^{J} I'_j$. Then, $\cup_{i \in I_1} \overline{B(i, r/2)}$ and $\cup_{i \in I_2} \overline{B(i, r/2)}$ are both path-connected. And $d(I_1, I_2) = r$ is also clear. This is so as follows. First, $d(I_1, I_2) \leqslant d(i_j, i_k) = r$. And second, for any small enough $\epsilon > 0$, $\cup_{i \in I_1} \overline{B(i, r/2 - \epsilon)}$ and $\cup_{i \in I_2} \overline{B(i, r/2 - \epsilon)}$ are not connected. Thus, $d(I_1, I_2) > r - 2\epsilon$. As ϵ can be arbitrarily small, $d(I_1, I_2) \geqslant r$. Hence, $d(I_1, I_2) = r$. □

Below is our main exponential inequality.

Theorem A.1. *Under Assumptions 1, A.1 and A.2, there exists some constant $C_a > 0$, which satisfies $C_a e^q C_d^q a_{cv}^{-dq} [d(q-1)]! \geqslant 1$ for any positive integer q, and*

$$P(|S_N| \geqslant N\epsilon) \leqslant \frac{\exp(2d+4)}{4\sqrt{\pi} d^{d-1/2}} \exp\left\{ -(d+1) \left[\frac{N\epsilon^2 (MeC_d a_{cv}^{-d} d^d)^{-2}}{16 \max(1,\, C_a + C_d^{-1} C_{cv} a_{cv}^{d-1} e^{2a_{cv}})} \right]^{\frac{1}{2d+2}} \right\}.$$

Proof of Theorem A.1: Before further discussion, let us define some notations. For $q \geqslant 2$, let $P_q \equiv \{\{i_1, i_2, \cdots, i_q\} \in \mathbb{N}^q : 1 \leqslant i_1 \leqslant i_2 \leqslant \cdots \leqslant i_q \leqslant N, \text{but they are not all equal}\}$. Thus, P_q is a collection of q natural numbers between 1 and N. For any $p_q = \{i_1, i_2, \cdots, i_q\} \in P_q$, by Lemma A.1, we can partition them into two non-empty mutually exclusive subsets $I_1(p_q)$ and $I_2(p_q)$, such that $I_1(p_q) \cup I_2(p_q) = p_q$, $d[I_1(p_q), I_2(p_q)] \equiv d[\{\vec{i} : i \in I_1(p_q)\}, \{\vec{i} : i \in I_2(p_q)\}] = r$, and both

$\cup_{i\in I_1(p_q)}\overline{B(\vec{i},r/2)}$ and $\cup_{i\in I_2(p_q)}\overline{B(\vec{i},r/2)}$ are path-connected. Then,

$$\begin{aligned}
&A_q(N) \equiv \sum_{1\leqslant i_1\leqslant\cdots i_q\leqslant N} |\operatorname{E} X_{i_1,N}\cdots X_{i_q,N}| \\
&\leqslant NM^q + \sum_{p_q\in P_q} |\operatorname{E}\prod_{j\in I_1(p_q)} X_{j,N}\cdot \operatorname{E}\prod_{j\in I_2(p_q)} X_{j,N}| + \sum_{p_q\in P_q} |\operatorname{cov}(\prod_{j\in I_1(p_q)} X_{j,N}, \prod_{j\in I_2(p_q)} X_{j,N})|,
\end{aligned} \tag{A.1}$$

where NM^q is an upper bound of the summation of all overlapped points, because when $i_1 = i_2 = \cdots = i_q$, $|\operatorname{E} X_i^q| \leqslant M^q$. The second term on the right hand side is bounded by

$$\sum_{p_q\in P_q} |\operatorname{E}\prod_{j\in I_1(p_q)} X_{j,N}\cdot \operatorname{E}\prod_{j\in I_2(p_q)} X_{j,N}| \leqslant \sum_{m=1}^{q-1} A_m(N)A_{q-m}(N). \tag{A.2}$$

It remains to evaluate the third term on the right hand side of Eq. (A.1). For any natural number $1 \leqslant i \leqslant N$, define $P_q(i) \equiv \{\{i_1, i_2, \cdots, i_q\} \in P_q : i = i_1\}$. Then $P_q = \cup_{i=1}^{N-1} P_q(i)$. Next, define $P_q(i, \lfloor r\rfloor) \equiv \{p_q \in P_q(i) : \lfloor r\rfloor \leqslant d(I_1(p_q), I_2(p_q)) < \lfloor r\rfloor + 1\}$, where $\lfloor r\rfloor$ is the largest integer that is not larger than r. Then

$$\begin{aligned}
&\sum_{p_q\in P_q} |\operatorname{cov}(\prod_{j\in I_1(p_q)} X_{j,N}, \prod_{j\in I_2(p_q)} X_{j,N})| = \sum_{i=1}^{N-1}\sum_{p_q\in P_q(i)} |\operatorname{cov}(\prod_{j\in I_1(p_q)} X_{j,N}, \prod_{j\in I_2(p_q)} X_{j,N})| \\
&\leqslant \sum_{i=1}^{N-1}\sum_{\lfloor r\rfloor=0}^{\infty}\sum_{p_q\in P_q(i,\lfloor r\rfloor)} |\operatorname{cov}(\prod_{j\in I_1(p_q)} X_{j,N}, \prod_{j\in I_2(p_q)} X_{j,N})|.
\end{aligned} \tag{A.3}$$

Here, once we fix the position i and consider a non-empty $P(i, \lfloor r\rfloor)$, we can sequentially establish a sequence of closed balls with radius r such that each ball contains at least one another point in $\{1 \leqslant i_1 \leqslant \cdots i_q \leqslant N\}$, so all points in $\{1 \leqslant i_1 \leqslant \cdots i_q \leqslant N\}$ can be covered sequentially in $(q-1)$ balls. In $\mathbb{R}^d$, the number of points in a ball of radius r with distance greater than or equal to 1, is less than or equal to $C_d(\lfloor r\rfloor + 1)^d$ under Assumption 1. Thus, when i is fixed,

$\sum_{p_q\in P_q(i,\lfloor r\rfloor)} 1 \leqslant \{C_d(\lfloor r\rfloor+1)^d\}^{q-1}$. Then by Assumption A.2 and Eq. (A.3),

$$\begin{aligned}&\sum_{p_q\in P_q}|\mathrm{cov}(\prod_{j\in I_1(p_q)}X_{j,N},\prod_{j\in I_2(p_q)}X_{j,N})|\leqslant\sum_{i=1}^{N-1}\sum_{\lfloor r\rfloor=0}^{\infty}\sum_{p_q\in P_q(i,\lfloor r\rfloor)}C_{cv}M^qe^qe^{-a_{cv}\lfloor r\rfloor}\\&\leqslant\sum_{i=1}^{N}\sum_{\lfloor r\rfloor=0}^{\infty}\{C_d(\lfloor r\rfloor+1)^d\}^{q-1}C_{cv}M^qe^qe^{-a_{cv}\lfloor r\rfloor}=C_d^{q-1}C_{cv}M^qe^qN\sum_{\lfloor r\rfloor=0}^{\infty}(\lfloor r\rfloor+1)^{d(q-1)}e^{-a_{cv}\lfloor r\rfloor}\\&\leqslant C_d^{q-1}C_{cv}M^qe^qN\sum_{k=0}^{\infty}\int_{k+1}^{k+2}x^{d(q-1)}e^{-a_{cv}(x-2)}dx\leqslant C_d^{q-1}C_{cv}M^qNe^{2a_{cv}+q}a_{cv}^{d-dq-1}[d(q-1)]!.\end{aligned}$$

From the Bohr–Mollerup theorem (Olver, 2010, p.138), $\ln\Gamma(x)$ is a convex function on $(0,\infty)$. Thus, $\ln\{C_d^{q-1}C_{cv}e^{2a_{cv}+q}a_{cv}^{d-dq-1}[d(q-1)]!\}=\ln(C_d^{-1}C_{cv}a_{cv}^{d-1}e^{2a_{cv}})+q\ln(C_dea_{cv}^{-d})+\ln\{[d(q-1)]!\}$ is a convex function with respect to q and it goes to infinity as $q\to\infty$. Accordingly, $C_d^qe^qa_{cv}^{-dq}[d(q-1)]!$ has a minimum value, denoted $1/C_a$, i.e., $C_ae^qC_d^qa_{cv}^{-dq}[d(q-1)]!\geqslant 1$. As a result,

$$\begin{aligned}&NM^q+\sum_{p_q\in P_q}|\mathrm{cov}(\prod_{j\in I_1(p_q)}X_{j,N},\prod_{j\in I_2(p_q)}X_{j,N})|\\&\leqslant(C_a+C_d^{-1}C_{cv}a_{cv}^{d-1}e^{2a_{cv}})NM^qe^qC_d^qa_{cv}^{-dq}[d(q-1)]!\equiv V_q(N).\end{aligned}$$

By the above inequality and Eq. (A.1) and (A.2), it follows $A_q(N)\leqslant\sum_{m=1}^{q-1}A_m(N)A_{q-m}(N)+V_q(N)$. For this inequality, Lemma 12 in Doukhan and Louhichi (1999) may be applicable to obtain a bound for $A_q(N)$, if $\max(V_2^{m/2},V_m)\max(V_2^{(q-m)/2},V_{q-m})\leqslant\max(V_2^{q/2},V_q)$ holds. By "a Technical Lemma" in Doukhan and Louhichi (1999, p. 336), it is sufficient that $V_q(N)$ satisfies the convexity condition: $V_p\leqslant V_q^{(p-2)/(q-2)}V_2^{(q-p)/(q-2)}$, which holds because $\ln V_q(N)$ is a convex function with respect to q. Hence Lemma 12 in Doukhan and Louhichi (1999) is applicable. Denote $E_N=\max(1,\,C_a+C_d^{-1}C_{ac}a_{cv}^{d-1}e^{2a_{cv}})N$ and $B=MeC_da_{cv}^{-d}$, then

$$\begin{aligned}&|\,\mathrm{E}\,S_N^q|\leqslant q!A_q(N)\leqslant q!\cdot\frac{1}{q}\binom{2q-2}{q-1}\max\{E_NB^q[d(q-1)]!,(E_NB^2d!)^{q/2}\}\\&\leqslant\frac{(2q-2)!}{(q-1)!}E_N^{q/2}B^q[d(q-1)]!,\end{aligned}$$

where the second inequality comes from $E_N\geqslant 1$ and $[d(q-1)]!=(1\cdot2\cdot\dots\cdot d)[(d+1)(d+2)\cdots(2d)]\cdots[(dq-2d+1)\cdots(d(q-1))]\geqslant(d!)^{q-1}\geqslant(d!)^{q/2}$.

From Stirling's formula, $\sqrt{2\pi n}(n/e)^n\leqslant n!\leqslant e\sqrt{n}(n/e)^n$ (Abramowitz and Stegun, 1967, p.

257). Let $q = 2p$, where $p \geqslant 1$. For any $\epsilon > 0$,

$$
\begin{aligned}
&P(|S_N| \geqslant N\epsilon) = P(|S_N^{2p}| \geqslant N^{2p}\epsilon^{2p}) \\
&\leqslant \mathrm{E}\, S_N^{2p}/(N\epsilon)^{2p} \leqslant \frac{(4p-2)!}{(2p-1)!} E_N^p B^{2p}[d(2p-1)]!/(N\epsilon)^{2p} \\
&\leqslant (\frac{E_N B^2}{N^2\epsilon^2})^p \frac{e(4p-2)^{1/2}((4p-2)/e)^{4p-2}}{(2\pi(2p-1))^{1/2}((2p-1)/e)^{2p-1}} \cdot e[d(2p-1)]^{1/2}[d(2p-1)/e]^{d(2p-1)} \\
&= \frac{\exp(d+3)}{4\sqrt{\pi}d^{d-1/2}}(\frac{16E_N B^2 d^{2d}}{N^2\epsilon^2 e^{2d+2}})^p (2p-1)^{(2p-1)(d+1)+1/2} \\
&\leqslant \frac{\exp(d+3)}{4\sqrt{\pi}d^{d-1/2}}(\frac{16E_N B^2 d^{2d}}{N^2\epsilon^2 e^{2d+2}})^p (2p)^{2p(d+1)} = \frac{\exp(d+3)}{4\sqrt{\pi}d^{d-1/2}}\left[(C_N q)^q\right]^{d+1},
\end{aligned}
$$

where $C_N^{2(d+1)} \equiv 16E_N B^2 d^{2d}/(N^2\epsilon^2 e^{2d+2})$. Since the upper bound depends on q, which is arbitrary, one may select q to minimize the upper bound. For that purpose, let $h(q) = q\ln C_N + q\ln q$. As $h'(q) = \ln C_N + \ln q + 1$ and $h''(q) = 1/q > 0$, $h(q)$ is a convex function. The q value which minimizes $h(q)$ is $q^* = e^{-1}C_N^{-1}$. However since $q \in \mathbb{Z}$, but q^* might not be an integer, so we pick $q_0 = \lfloor q^* \rfloor = \lfloor e^{-1}C_N^{-1} \rfloor$. Thus,

$$
(C_N q_0)^{q_0} = (C_N \lfloor e^{-1}C_N^{-1} \rfloor)^{\lfloor e^{-1}C_N^{-1} \rfloor} \leqslant (e^{-1})^{\lfloor e^{-1}C_N^{-1} \rfloor} < (e^{-1})^{e^{-1}C_N^{-1}-1} = \exp(1 - e^{-1}C_N^{-1}).
$$

Recall $E_N = \max(1,\, C_a + C_d^{-1}C_{cv}a_{cv}^{d-1}e^{2a_{cv}})N$. Therefore,

$$
\begin{aligned}
&P(|S_N| \geqslant N\epsilon) \leqslant \frac{\exp(d+3)}{4\sqrt{\pi}d^{d-1/2}} \exp(d+1-(d+1)e^{-1}C_N^{-1}) \\
&\leqslant \frac{\exp(d+3)}{4\sqrt{\pi}d^{d-1/2}} \exp\left[d+1-(d+1)e^{-1}\left(\frac{N^2\epsilon^2 e^{2d+2}}{16E_N B^2 d^{2d}}\right)^{1/(2d+2)}\right] \\
&= \frac{\exp(2d+4)}{4\sqrt{\pi}d^{d-1/2}} \exp\left\{-(d+1)\left[\frac{N\epsilon^2}{16\max(1,\, C_a + C_d^{-1}C_{cv}a_{cv}^{d-1}e^{2a_{cv}})(MeC_d a_{cv}^{-d}d^d)^2}\right]^{1/(2d+2)}\right\}.
\end{aligned}
$$

□

When $d = 1$, we have the same decreasing rate $\exp(-\,\mathrm{const}\cdot(N\epsilon^2)^{1/4})$ as that in Doukhan and Louhichi (1999) for weakly dependent time series. But in higher dimensional space, the convergence rate is slower. For independent random variables sequence in the Bernstein inequality, the rate is $\exp(-\,\mathrm{const}\cdot N\epsilon^2)$, so the decreasing rates for random variables with spatial dependence can be slower.

Theorem A.1 can be applied to mixing and NED random fields by exploring their implied covariance structures for Assumption A.2. When $\{X_{i,N}\}_{i=1}^{N}$ is a centered α-mixing random field bounded by M and its α-mixing coefficients satisfies $\alpha(u,v,r) \leqslant (u+v)^{\tau} e^{-a_{cv} r}$ for some constant $a_{cv} > 0$. Then, from Lemma 3 in Doukhan (1994, p.10), $|\text{cov}(X_{i_1,N} \cdots X_{i_u,N}, X_{j_1,N} \cdots X_{j_v,N})| \leqslant 4M^q q^{\tau} e^{-a_{cv} r}$. As $4q^{\tau} \leqslant C_{\tau} e^q$ for all integers $q \geqslant 0$, for some $C_{\tau} > 0$, Assumption A.2 is satisfied and Theorem A.1 is applicable:

Corollary A.1. *Assume that $\{X_{i,N} : \vec{i} \in D_N\}_{N=1}^{\infty}$ is an α-mixing random field with α-mixing coefficient $\alpha(u,v,r) \leqslant (u+v)^{\tau} e^{-a_{cv} r}$ for some constant $a_{cv} > 0$ and it satisfies Assumptions 1 and A.1, then, for some constant $C_a > 0$ that depends only on d and a_{cv},*

$$P(|S_N| \geqslant N\epsilon) \leqslant \frac{\exp(2d+4)}{4\sqrt{\pi} d^{d-1/2}} \exp\left\{ -(d+1) \left(\frac{N\epsilon^2 (MeC_d a_{cv}^{-d} d^d)^{-2}}{16 \max(1, C_a + C_d^{-1} C_{\tau} a_{cv}^{d-1} e^{2a_{cv}})} \right)^{\frac{1}{2d+2}} \right\},$$

where C_{τ} is a constant satisfying $4q^{\tau} \leqslant C_{\tau} e^q$ for all integers $q > 0$.

Consider $\{X_{i,N}\}_{i=1}^{N}$ being a centered L_2-NED random field on an α-mixing random field $\{\epsilon_{i,N}\}_{i=1}^{N}$ as its base. Denote $\mathcal{F}_{i,n}(s) \equiv \sigma(\{\epsilon_{j,n} : d(\vec{j}, \vec{i}) \leqslant s\})$. Before further discussion, we need to obtain some covariance inequalities for NED random fields.

Lemma A.2. *Let $\epsilon_N = \{\epsilon_{i,N}, \vec{i} \in D_N, N \geqslant 1\}$ be an α-mixing random field with α-mixing coefficient $\alpha(u,v,r)$. $Z = \{Z_{i,N}, \vec{i} \in D_N, N \geqslant 1\}$ is an L_2-NED random field on ϵ_N such that $||Z_{i,N} - \mathrm{E}(Z_{i,N}|\mathcal{F}_{i,N}(s))||_{L^2} \leqslant d_{i,N}\psi(s)$. $f(x_1, \cdots, x_u)$ and $g(x_1, \cdots, x_v)$ are two bounded and Lipschitz functions, with bounds b_f and b_g and Lipschitz coefficients $Lip(f)$ and $Lip(g)$. Denote $r = \min_{m,l}\{d(\vec{i}_m, \vec{j}_l) : 1 \leqslant m \leqslant u, 1 \leqslant l \leqslant v\}$. When $r > 0$, for any positive $s < r/2$,*

$$|\text{cov}(f(Z_{i_1,N}, \cdots, Z_{i_u,N}), g(Z_{j_1,N}, \cdots, Z_{j_v,N}))| \leqslant [Lip(f) \sum_{k=1}^{u} d_{i_k,N}][Lip(g) \sum_{k=1}^{v} d_{j_k,N}]\psi^2(s)$$
$$+ 2b_g Lip(f) \sum_{k=1}^{u} d_{i_k,N}\psi(s) + 2b_f Lip(g) \sum_{k=1}^{v} d_{j_k,N}\psi(s) + 4b_f b_g \alpha(u,v,r-2s).$$

Proof of Lemma A.2: Let $\mathcal{F}_N(s) = \sigma(\cup_{j=1}^{u} \mathcal{F}_{i_j,N}(s))$, $f^s = \mathrm{E}[f(Z_{i_1,N}, \cdots, Z_{i_u,N})|\mathcal{F}_N(s)]$,

$\Delta f = f - f^s$. Similarly, define g^s and Δg. Then

$$
\begin{aligned}
&||\Delta f||_{L^2} \leqslant ||f(Z_{i_1,N}, \cdots, Z_{i_u,N}) - f(\mathrm{E}[Z_{i_1,N}|\mathcal{F}_{i_1,N}(s)], \cdots, \mathrm{E}[Z_{i_u,N}|\mathcal{F}_{i_u,N}(s)])||_{L^2} \\
\leqslant & Lip(f) \sum_{k=1}^{u} ||Z_{i_k,N} - \mathrm{E}[Z_{i_k,N}|\mathcal{F}_{i_k,N}(s)]||_{L^2} \leqslant [Lip(f) \sum_{k=1}^{u} d_{i_k,N}]\psi(s).
\end{aligned}
$$

As ϵ_n is an α-mixing random field, by Lemma 3 in Doukhan (1994, p.10), $|\mathrm{cov}(f^s,\ g^s)| \leqslant 4b_f b_g \alpha(u, v, r - 2s)$. Besides, because $\mathrm{E}\,\Delta f = 0$, $|\mathrm{cov}(\Delta f,\ g^s)| = |\int(\Delta f - \mathrm{E}\Delta f) \cdot (g^s - \mathrm{E}g^s)dP| \leqslant \int |\Delta f| \cdot |g^s - \mathrm{E}g^s|dP \leqslant 2b_g \int |\Delta f|dP \leqslant 2b_g||\Delta f||_2$ by Lyapunov's inequality. So,

$$
\begin{aligned}
&|\mathrm{cov}(f(Z_{i_1,N}, \cdots, Z_{i_u,N}), g(Z_{i_1,N}, \cdots, Z_{i_v,N}))| \\
\leqslant & |\mathrm{cov}(\Delta f,\ \Delta g)| + |\mathrm{cov}(\Delta f,\ g^s)| + |\mathrm{cov}(f^s,\ \Delta g)| + |\mathrm{cov}(f^s,\ g^s)| \\
\leqslant & ||\Delta f||_2||\Delta g||_{L^2} + 2b_g||\Delta f||_{L^2} + 2b_f||\Delta g||_{L^2} + 4b_f b_g \alpha(u, v, r - 2s) \\
\leqslant & [Lip(f) \sum_{k=1}^{u} d_{i_k,N}][Lip(g) \sum_{k=1}^{v} d_{j_k,N}]\psi^2(s) + 2b_g Lip(f) \sum_{k=1}^{u} d_{i_k,N}\psi(s) \\
& + 2b_f Lip(g) \sum_{k=1}^{v} d_{j_k,N}\psi(s) + 4b_f b_g \alpha(u, v, r - 2s).
\end{aligned}
$$

□

The inequality above can be applied to $f(x_1, \cdots, x_u) = \prod_{j=1}^{u} x_j$ and $g(x_1, \cdots, x_v) = \prod_{j=1}^{v} x_j$. Under Assumption A.1, $b_f = M^u$, $b_g = M^v$, $Lip(f) = M^{u-1}$ and $Lip(g) = M^{v-1}$. By taking $s = r/3$ in Lemma A.2 as done in the proof of Lemma A.3 in JP (2012), we have

Corollary A.2. *Under Assumptions 1, let $\epsilon_N = \{\epsilon_{i,N}, \vec{i} \in D_N\}$ be an α-mixing random field with α-mixing coefficients $\alpha(u, v, r) \leqslant (u+v)^\tau e^{-a_\epsilon r}$ for some $\tau > 0$ and $a_\epsilon > 0$. Let $\{X_{i,N} : \vec{i} \in D_N\}$ be an L_2-NED random field on ϵ_N satisfying Assumption A.1: $||X_{i,N} - E[X_{i,N}|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_X e^{-a_X s}$ for some $C_X > 0$ and $a_X > 0$. Denote $q \equiv u + v$ and $r = d(\{\vec{i}_1, \cdots, \vec{i}_u\}, \{\vec{j}_1, \cdots, \vec{j}_v\})$. Then*

$$
|\mathrm{cov}(X_{i_1,N} \cdots X_{i_u,N}, X_{j_1,N} \cdots X_{j_v,N})| \leqslant (\frac{q^2 C_X^2}{4M^2} + \frac{2qC_X}{M} + 4q^\tau)M^q \exp[-\frac{r \min(a_X, a_\epsilon)}{3}].
$$

Combining Theorem A.1 and Corollary A.2, the exponential inequality of L_2-NED random fields follows:

Corollary A.3. *Under Assumptions 1 and A.1, let $\{X_{i,N} : \vec{i} \in D_N\}$ and $\epsilon_N = \{\epsilon_{i,N} : \vec{i} \in D_N\}$*

satisfy the conditions in Corollary A.2. Define $C_{\tau MC} \equiv \sup_{0<q\in\mathbb{Z}}(\frac{q^2C_X^2}{4M^2} + \frac{2qC_X}{M} + 4q^\tau)e^{-q}$*. Then for some constant* $C_a > 0$ *that depends only on* d *and* $a_{cv} \equiv \min(a_X, a_\epsilon)/3$*,*

$$P(|S_N| \geqslant N\epsilon) \leqslant \frac{\exp(2d+4)}{4\sqrt{\pi}d^{d-1/2}} \exp\left\{-(d+1)\left(\frac{N\epsilon^2(MeC_d a_{cv}^{-d} d^d)^{-2}}{16\max(1, C_a + C_d^{-1}C_{\tau MC}a_{cv}^{d-1}e^{2a_{cv}})}\right)^{1/(2d+2)}\right\}.$$

Corollary A.3 can further be generalized to unbounded NED random fields having a uniform exponential bound.

Assumption A.3. $C_{EB} = \sup_{i,N} \mathrm{E}\exp(\gamma|X_{i,N}|^\alpha) < \infty$ *for some* $\alpha > 0$ *and* $\gamma > 0$.

Assumptions A.3 implies that all orders of moments of $X_{i,N}$ exist and $\sup_{i,N} P(|X_{i,N}| > M) \leqslant C_{EB}\exp(-\gamma M^\alpha)$.

Theorem A.2. *Under Assumption 1, let* $\epsilon_N = \{\epsilon_{i,N}, \vec{i} \in D_N\}$ *be an* α*-mixing random field with* α*-mixing coefficients* $\alpha(u,v,r) \leqslant (u+v)^\tau e^{-a_\epsilon r}$ *for some* $\tau > 0$ *and* $a_\epsilon > 0$*. Let* $\{X_{i,N} : \vec{i} \in D_N\}$ *be an* L_2*-NED random field on* ϵ_N *satisfying Assumption A.3 such that* $||X_{i,N} - \mathrm{E}[X_{i,N}|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_X e^{-a_X s}$ *for some* $C_X > 0$ *and* $a_X > 0$*. Then, for some positive constants* C_a *and* $C_{\tau C}$*,*

$$\begin{aligned} P(|S_N| \geqslant N\epsilon) \leqslant &\left[4(\sup_{i,N}||X_{i,N}||_{L^p})\epsilon^{-1}C_{EB}^{1/q} + \frac{\exp(2d+4)}{4\sqrt{\pi}d^{d-1/2}}\right] \\ &\exp\left\{-\left[\frac{N\epsilon^2(d+1)^{2d+2}(\gamma/q)^{2/\alpha}}{64\max(1, C_a + C_d^{-1}C_{\tau C}a_{cv}^{d-1}e^{2a_{cv}})(eC_d a_{cv}^{-d}d^d)^2}\right]^{\alpha/[(2d+2)\alpha+2]}\right\}, \end{aligned}$$

where $p > 0$ *and* $q > 0$ *satisfy* $p^{-1} + q^{-1} = 1$ *and* $a_{cv} \equiv \min(a_X, a_\epsilon)/3$.

Proof of Theorem A.2: For any $M \geqslant 1$, define $f_M(x) \equiv x1(|x| \leqslant M) + M1(x > M) - M1(x < -M)$. Apparently, $f_M(x)$ is weakly increasing, $|f_M(x)| \leqslant M$ for all $x \in \mathbb{R}$, and $|f_M(X_{i,N}) - \mathrm{E}\, f_M(X_{i,N})| \leqslant 2M$. Notice that $|f_M(x_1) - f_M(x_2)| \leqslant |x_1 - x_2|$ for all x_1, $x_2 \in \mathbb{R}$. $\{\overline{X_{i,N}} \equiv f_M(X_{i,N})\}_{i=1}^N$ is a bounded NED random field with $||\overline{X_{i,N}} - \mathrm{E}[\overline{X_{i,N}}|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_X e^{-a_X s}$. Define $\widetilde{X_{i,N}} = X_{i,N} - \overline{X_{i,N}}$. From Corollary A.3, for any $\epsilon > 0$,

$$
\begin{aligned}
&P(|S_N| \geqslant N\epsilon) \leqslant P(|\sum_{i=1}^{N}(\overline{X_{i,N}} - \mathrm{E}\,\overline{X_{i,N}})| \geqslant \frac{N\epsilon}{2}) + P(|\sum_{i=1}^{N}(\widetilde{X_{i,N}} - \mathrm{E}\,\widetilde{X_{i,N}})| \geqslant \frac{N\epsilon}{2})\\
\leqslant&\frac{\exp(2d+4)}{4\sqrt{\pi}d^{d-1/2}}\exp\left\{-(d+1)\left(\frac{N\epsilon^2}{16\max(1, C_a + C_d^{-1}C_{\tau C}a_{cv}^{d-1}e^{2a_{cv}})(2MeC_d a_{cv}^{-d}d^d)^2}\right)^{1/(2d+2)}\right\}\\
&+P(|\sum_{i=1}^{N}(\widetilde{X_{i,N}} - E\widetilde{X_{i,N}})| \geqslant \frac{N\epsilon}{2}),
\end{aligned}
$$

where $C_{\tau C}$ satisfies $\frac{q^2C_X^2}{4M^2} + \frac{2qC_X}{M} + 4q^\tau \leqslant \frac{q^2C_X^2}{4} + 2qC_X + 4q^\tau \leqslant C_{\tau C}e^q$ for all integers $q > 0$. Consider $q = 1$ and we obtain $C_{\tau C} \geqslant 4e^{-1} > 1$.

Notice that $|\widetilde{X_{i,N}}| \leqslant |X_{i,N}|1(|X_{i,N}| \geqslant M)$. Thus, $\mathrm{E}\,|\widetilde{X_{i,N}}| \leqslant \mathrm{E}[|X_{i,N}|1(|X_{i,N}| \geqslant M)] \leqslant ||X_{i,N}||_{L^p}||1(|X_{i,N}| \geqslant M)||_{L^q} \leqslant ||X_{i,N}||_{L^p}C_{EB}^{1/q}\exp(-\frac{\gamma}{q}M^\alpha)$ by Hölder's inequality, where $p > 1$ and $p^{-1} + q^{-1} = 1$. Notice that we can have an infinity combinations of (p, q). Thus, we have

$$
\begin{aligned}
&P(|\sum_{i=1}^{N}(\widetilde{X_{i,N}} - \mathrm{E}\,\widetilde{X_{i,N}})| \geqslant \frac{N\epsilon}{2}) \leqslant \mathrm{E}\,|\sum_{i=1}^{N}(\widetilde{X_{i,N}} - \mathrm{E}\,\widetilde{X_{i,N}})|/(N\epsilon/2)\\
\leqslant&4\sum_{i=1}^{N}\mathrm{E}\,|\widetilde{X_{i,N}}|/(N\epsilon) \leqslant 4(\sup_{i,N}||X_{i,N}||_{L^p})\epsilon^{-1}C_{EB}^{1/q}\exp(-\frac{\gamma}{q}M^\alpha).
\end{aligned}
\tag{A.4}
$$

Because $C_{\tau C} > 1$, with Eq. (A.4), we have

$$
\begin{aligned}
&P(|S_N| \geqslant N\epsilon) \leqslant 4(\sup_{i,N}||X_{i,N}||_{L^p})\epsilon^{-1}C_{EB}^{1/q}\exp(-\frac{\gamma}{q}M^\alpha)+\\
&\frac{\exp(2d+4)}{4\sqrt{\pi}d^{d-1/2}}\exp\left\{-(d+1)\left(\frac{N\epsilon^2}{16\max(1, C_a + C_d^{-1}C_{\tau C}a_{cv}^{d-1}e^{2a_{cv}})(2MeC_d a_{cv}^{-d}d^d)^2}\right)^{\frac{1}{2d+2}}\right\},
\end{aligned}
$$

By taking M such that the rates of the two exponential functions are the same, i.e.,

$$
\frac{\gamma}{q}M^\alpha = (d+1)\left(\frac{N\epsilon^2}{16\max(1, C_a + C_d^{-1}C_{\tau C}a_{cv}^{d-1}e^{2a_{cv}})(2MeC_d a_{cv}^{-d}d^d)^2}\right)^{\frac{1}{2d+2}},
$$

we have

$$
M = \left(\frac{N\epsilon^2[q(d+1)/\gamma]^{2d+2}}{16\max(1, C_a + C_d^{-1}C_{\tau C}a_{cv}^{d-1}e^{2a_{cv}})(2eC_d a_{cv}^{-d}d^d)^2}\right)^{\frac{1}{(2d+2)\alpha+2}}.
$$

Therefore,

$$P(|S_N| \geqslant N\epsilon) \leqslant \left[4(\sup_{i,N} ||X_{i,N}||_{L^p})\epsilon^{-1}C_{EB}^{1/q} + \frac{\exp(2d+4)}{4\sqrt{\pi}d^{d-1/2}}\right]$$
$$\exp\left\{-\left[\frac{N\epsilon^2(d+1)^{2d+2}(\gamma/q)^{2/\alpha}}{64\max(1, C_a + C_d^{-1}C_{\tau C}a_{cv}^{d-1}e^{2a_{cv}})(eC_d a_{cv}^{-d}d^d)^2}\right]^{\alpha/[(2d+2)\alpha+2]}\right\}.$$

□

For an unbounded random field, $P(|S_N| \geqslant N\epsilon) \leqslant \text{const} \cdot \exp[-\text{const} \cdot N^{1/(2d+4)}]$ if $\alpha = 1$, a slightly slower decreasing rate than that of a bounded NED random field, $\exp[-\text{const} \cdot N^{1/(2d+2)}]$. But if we have stronger conditions in Assumption A.3, i.e., larger α, we have a faster decreasing rate. As $\alpha \to \infty$, the limit decreasing rate will be exactly that for the bounded one.

B. Some Properties of $h(u|\delta)$ and $H(u|\delta)$

For $\delta = (\delta_1, \delta_2, \cdots)$, denote $\psi(u|\delta) \equiv 1 + \sum_{l=1}^{\infty} \delta_l \sqrt{2}\cos l\pi u$. Then, $h(u|\delta) = (1-\epsilon_0)\psi^2(u|\delta)/(1 + \sum_{k=1}^{\infty}\delta_k^2) + \epsilon_0$. Clearly, $||\delta||_k \equiv \sum_{i=1}^{\infty} i^k|\delta_i|$ is nondecreasing in k. In particular, when $k = 0$, $||\delta||_0 = \sum_{i=1}^{\infty}|\delta_i|$. $||\delta||_k < \infty$ implies most entries in the tail of $(\delta_1, \delta_2, \delta_3, \cdots)$ are small. When k is large, the frequency of $\cos k\pi u$ is high. Thus, $||\delta||_k < \infty$ limits the effect of high frequency basis functions. Denote $\psi^{(0)}(u|\delta) = \psi(u|\delta)$, $\psi^{(m)}(u|\delta) = \partial^m\psi(u|\delta)/\partial u^m$ for $m = 1, 2, \cdots$, and $\nabla_{\delta_j} \equiv \partial/\partial\delta_j$. Similarly, $h^{(m)}(u|\delta)$'s are defined. Since $d^k \cos u/d^k u = \cos(u + k\pi/2)$, $\psi^{(m)}(u|\delta) = 1(m=0) + \sum_{k=1}^{\infty}\delta_k\sqrt{2}(k\pi)^m\cos(k\pi u + m\pi/2)$. Thus, $|\psi^{(m)}(u|\delta)| \leqslant 1(m=0) + \sqrt{2}\pi^m||\delta||_m$. Because $h(u|\delta)$, $H(u|\delta)$, $\psi_1(u|\delta) \equiv \frac{u}{H(u|\delta)}$ and their derivatives appear in the first and second order derivatives of the log-likelihood function, properties of these terms are used in proofs. We summarize them in this section.

Lemma B.1. *Let $0 \leqslant m \in \mathbb{Z}$. Denote $\delta^1 = (\delta_{11}, \delta_{12}, \cdots)$ and $\delta^2 = (\delta_{21}, \delta_{22}, \cdots)$.*

(1) $\sup_{u\in[0,1]}|h^{(m)}(u|\delta)| \leqslant 2^m(1 + \sqrt{2}\pi^m||\delta||_m)^2$.

(2) $\sup_{u\in[0,1]}|\nabla_{\delta_j}h^{(m)}(u|\delta)| < (1 + \sqrt{2}\pi^m||\delta||_m)^2 2^{m+2}\pi^m j^m$.

(3) $\sup_{u\in[0,1]}|h^{(m)}(u|\delta^1) - h^{(m)}(u|\delta^2)| \leqslant [2^m||\delta^1 + \delta^2||_0(1 + \sqrt{2}\pi^m||\delta^1||_m)^2 + 2\sqrt{2}(1+\pi)^m + 2^{m+1}\pi^m(||\delta^1||_m + ||\delta^2||_m)] \cdot ||\delta^1 - \delta^2||_m$.

(4) $\sup_{u\in[0,1]}|\nabla_{\delta_j}h(u|\delta^1) - \nabla_{\delta_j}h(u|\delta^2)| = [4(||\delta^1||_0 + ||\delta^2||_0 + 1)(\sqrt{2} + ||\delta^1||_0)^2 + 1] \cdot ||\delta^1 - \delta^2||_0$. *When $m \geqslant 1$,* $|\nabla_{\delta_j}h^{(m)}(u|\delta^1) - \nabla_{\delta_j}h^{(m)}(u|\delta^2)| \leqslant \pi^m\{4(1+j)^m + [2^{1.5}j^m + 4(1+j)^m||\delta^2||_m + 2^{m+1}] \cdot$

$(||\delta^1||_m + ||\delta^2||_m) + 2\sqrt{2}(1+\pi^{-1})^m + (2 + 2||\delta^1+\delta^2||_0)(\frac{2}{\pi})^m(1+\sqrt{2}\pi^m||\delta^2||_m)^2\} \cdot ||\delta^1 - \delta^2||_m$.

(5) $\sup_{u\in[0,1]} |\nabla_{\delta_i,\delta_j} h(u|\delta)| < 4 + 2[1(i=j)+4](1+\sqrt{2}||\delta||_0)^2$. $\sup_{u\in[0,1]} |\nabla_{\delta_i,\delta_j} h^{(m)}(u|\delta)| \leqslant [1+(1+\sqrt{2}\pi^m||\delta||_m)^2]2^{m+1}(1+2\pi^m j^m + 2\pi^m i^m)$ *for* $m \geqslant 1$.

(6) $\sup_{u\in[0,1]} |\nabla_{\delta_i,\delta_j} h(u|\delta^1) - \nabla_{\delta_i,\delta_j} h(u|\delta^2)| \leqslant 4||\delta^1-\delta^2||_0 \cdot (2 + ||\delta^1||_0 + ||\delta^2||_0)[3 + 2\sqrt{2}||\delta^1||_0 + 4(1+\sqrt{2}||\delta^1||_0)^2]$.

(7) $\sup_{u\in[0,1]} |h'(u|\delta)/h(u|\delta)| \leqslant \pi||\delta||_1(2/\epsilon_0)^{1/2}$.

(8) $|\frac{1}{h(u|\delta^1)} - \frac{1}{h(u|\delta^2)}| \leqslant \epsilon_0^{-2}[||\delta^1+\delta^2||_0(1+\sqrt{2}||\delta^1||_0)^2 + 2\sqrt{2} + 2(||\delta^1||_0 + ||\delta^2||_0)]||\delta^1-\delta^2||_0$.

Lemma B.2. (1) $H(u|\delta)/u \geqslant \epsilon_0$.

(2) $[H(u|\delta_1) - H(u|\delta_2)]/u \leqslant \sup_{v\in[0,1]} |h(v|\delta_1) - h(v|\delta_2)|$.

(3) $\nabla_{\delta_k} H(u|\delta) = \frac{2(1-\epsilon_0)}{1+\sum_{j=1}^\infty \delta_j^2}\{\sqrt{2}\frac{\sin k\pi u}{k\pi} + \sum_{j=1}^\infty \delta_j \frac{\sin(k+j)\pi u}{(k+j)\pi} + \sum_{j\neq k}\delta_j\frac{\sin(k-j)\pi u}{(k-j)\pi} - \frac{\delta_k[H(u|\delta)-u]}{1-\epsilon_0}\}$.

(4) $\sup_{u\in(0,1)} |\nabla_{\delta_k} H(u|\delta)/u| \leqslant 4(1+\sqrt{2}||\delta||_0)^2$.

(5) $\sup_{u\in(0,1)} |\nabla_{\delta_k} H(u|\delta)/H(u|\delta)| \leqslant 1 + 2\epsilon_0^{-1}(\sqrt{2} + 2||\delta||_0)$.

(6) $\sup_{u\in(0,1)} |\nabla_{\delta_k} H(u|\delta^1) - \nabla_{\delta_k} H(u|\delta^2)|/u \leqslant [4(||\delta^1||_0 + ||\delta^2||_0 + 1)(\sqrt{2} + ||\delta^1||_0)^2 + 1] \cdot ||\delta^1 - \delta^2||_0$.

(7) $\sup_{u\in(0,1)} |\nabla_{\delta_k,\delta_j} H(u|\delta)/u| \leqslant 4 + 10(1+\sqrt{2}||\delta||_0)^2$.

(8) $\sup_{u\in(0,1)} |\nabla_{\delta_k,\delta_j} H(u|\delta^1) - \nabla_{\delta_k,\delta_j} H(u|\delta^2)|/u \leqslant 4(1 + ||\delta^1||_0 + ||\delta^2||_0)[5 + 2\sqrt{2}||\delta^1||_0 + 3(1+\sqrt{2}||\delta^1||_0)^2] \cdot ||\delta^1-\delta^2||_0$.

(9) $\sup_{u\in(0,1)} |u\partial[\nabla_{\delta_k} H(u|\delta)/u]/\partial u| \leqslant 8(1+\sqrt{2}||\delta||_0)^2$.

(10) $\sup_{u\in(0,1)} |u\partial[\nabla_{\delta_k} H(u|\delta)/H(u|\delta)]/\partial u| \leqslant C(1+\sqrt{2}||\delta||_0)^4$ *for some constant* $C > 0$.

(11) $\sup_{u\in(0,1)} |u\partial[\nabla_{\delta_k,\delta_j} H(u|\delta)/H(u|\delta)]/\partial u| \leqslant C(1+\sqrt{2}||\delta||_0)^4$ *for some constant* $C > 0$.

Lemma B.3. *Let* $\psi_1(u|\delta) = \frac{u}{H(u|\delta)}$.

(1) $\sup_{0<u<1} |\psi_1'(u|\delta)u| \leqslant C(1+\sqrt{2}||\delta||_0)^2$ *for some constant* $C > 0$.

(2) $\sup_{0<u<1} |\nabla_{\delta_j}\psi_1(u|\delta)| \leqslant 4\epsilon_0^{-2}(1+\sqrt{2}||\delta||_0)^2$.

(3) $\sup_{0<u<1} |\psi_1(u|\delta^1) - \psi_1(u|\delta^2)| \leqslant \epsilon_0^{-2}[(||\delta^1||_0 + ||\delta^2||_0)((1+\sqrt{2}||\delta^1||_0)^2 + 2) + 2\sqrt{2}] \cdot ||\delta^1 - \delta^2||_0$.

(4) $\sup_{0<u<1} |\psi_1'(u|\delta^1) - \psi_1'(u|\delta^2)|u \leqslant 2\epsilon_0^{-3}\left[\epsilon_0 + (1+\sqrt{2}||\delta||_0)^2\right]\{(||\delta^1||_0 + ||\delta^2||_0)[(1+\sqrt{2}||\delta||_0)^2 + 2] + 2\sqrt{2}\} \cdot ||\delta^1 - \delta^2||_0$.

(5) $\sup_{0<u<1} |\nabla_{\delta_j}\psi_1(u|\delta^1) - \nabla_{\delta_j}\psi_1(u|\delta^2)| \leqslant C(1 + ||\delta^2||_0)^2(||\delta^1||_0 + ||\delta^2||_0 + 1)(1 + ||\delta^1||_0)^2 \cdot ||\delta^2 - \delta^1||_0$ *for some constant* $C > 0$.

Proof of Lemma B.1: (1) By the Leibniz rule,

$$
\begin{aligned}
&|h^{(m)}(u|\delta)| = |\epsilon_0 1(m=0) + \sum_{k=0}^{m} \frac{m!}{k!(m-k)!}\psi^{(k)}(u|\delta)\psi^{(m-k)}(u|\delta)\frac{1-\epsilon_0}{1+\sum_{j=1}^{\infty}\delta_j^2}| \\
\leqslant&\epsilon_0 1(m=0) + (1-\epsilon_0)\sum_{k=0}^{m}\frac{m!}{k!(m-k)!}(1+\sqrt{2}\pi^k||\delta||_k)(1+\sqrt{2}\pi^{m-k}||\delta||_{m-k}) \\
\leqslant&\epsilon_0 1(m=0) + (1-\epsilon_0)\sum_{k=0}^{m}\frac{m!}{k!(m-k)!}(1+\sqrt{2}\pi^m||\delta||_m)^2 \\
=&\epsilon_0 1(m=0) + (1-\epsilon_0)2^m(1+\sqrt{2}\pi^m||\delta||_m)^2 \leqslant 2^m(1+\sqrt{2}\pi^m||\delta||_m)^2.
\end{aligned}
$$

(2) $(1+\sum_{k=1}^{\infty}\delta_k^2)[h(u|\delta)-\epsilon_0] = (1-\epsilon_0)\psi^2(u|\delta)$ implies

$$
(1+\sum_{k=1}^{\infty}\delta_k^2)\nabla_{\delta_j}h(u|\delta) + 2\delta_j[h(u|\delta)-\epsilon_0] = 2\sqrt{2}(1-\epsilon_0)\psi(u|\delta)\cos j\pi u. \tag{B.1}
$$

Then

$$
\begin{aligned}
&\left|\nabla_{\delta_j}h(u|\delta)\right| = \left|\frac{2\sqrt{2}(1-\epsilon_0)\psi(u|\delta)\cos j\pi u - 2\delta_j[h(u|\delta)-\epsilon_0]}{1+\sum_{k=1}^{\infty}\delta_k^2}\right| \\
\leqslant&2\sqrt{2}(1+\sqrt{2}||\delta||_0) + |h(u|\delta)-\epsilon_0| \leqslant (1+\sqrt{2}||\delta||_0)^2 + 2\sqrt{2}(1+\sqrt{2}||\delta||_0) < 4(1+\sqrt{2}||\delta||_0)^2.
\end{aligned}
$$

Because $h^{(m)}(u|\delta) = \sum_{k=0}^{m}\frac{m!}{k!(m-k)!}\psi^{(k)}(u|\delta)\psi^{(m-k)}(u|\delta)\frac{1-\epsilon_0}{1+\sum_{k=1}^{\infty}\delta_k^2}$ when $m \geqslant 1$,

$$
\begin{aligned}
&\nabla_{\delta_j}\left[(1+\sum_{k=1}^{\infty}\delta_k^2)h^{(m)}(u|\delta)\right] = \nabla_{\delta_j}\left[(1-\epsilon_0)\sum_{k=0}^{m}\frac{m!}{k!(m-k)!}\psi^{(k)}(u|\delta)\psi^{(m-k)}(u|\delta)\right] \\
=&(1-\epsilon_0)\sum_{k=0}^{m}\frac{m!}{k!(m-k)!}\left[\nabla_{\delta_j}\psi^{(k)}(u|\delta)\psi^{(m-k)}(u|\delta) + \psi^{(k)}(u|\delta)\nabla_{\delta_j}\psi^{(m-k)}(u|\delta)\right] \qquad \text{(B.2)} \\
=&2(1-\epsilon_0)\sum_{k=0}^{m}\frac{m!}{k!(m-k)!}\psi^{(m-k)}(u|\delta)\nabla_{\delta_j}\psi^{(k)}(u|\delta).
\end{aligned}
$$

Thus,

$$
\nabla_{\delta_j}h^{(m)}(u|\delta) = \frac{2\sqrt{2}(1-\epsilon_0)\sum_{k=0}^{m}\frac{m!}{k!(m-k)!}(j\pi)^k\cos(j\pi u + \frac{k\pi}{2})\psi^{(m-k)}(u|\delta) - 2\delta_j h^{(m)}(u|\delta)}{1+\sum_{k=1}^{\infty}\delta_k^2}. \tag{B.3}
$$

Then, conclusion (1) in this lemma implies

$$\left|\nabla_{\delta_j} h^{(m)}(u|\delta)\right| \leqslant 2\sqrt{2}\sum_{k=0}^{m}\frac{m!}{k!(m-k)!}j^k\pi^k(1+\sqrt{2}\pi^{m-k}||\delta||_{m-k}) + 2^m(1+\sqrt{2}\pi^m||\delta||_m)^2$$
$$\leqslant (1+\sqrt{2}\pi^m||\delta||_m)^2\left[2\sqrt{2}(1+j\pi)^m + 2^m\right] < (1+\sqrt{2}\pi^m||\delta||_m)^2 2^{m+2}\pi^m j^m,$$

where the last inequality is by $2\sqrt{2}(1+j\pi)^m + 2^m \leqslant 2\sqrt{2}(2j\pi)^m$.

(3) Because $\sup_{0\leqslant u\leqslant 1}|\psi^{(m)}(u|\delta^1) - \psi^{(m)}(u|\delta^2)| \leqslant \sqrt{2}\pi^m||\delta^1-\delta^2||_m$, it follows

$$|\psi^{(m-k)}(u|\delta^1)\psi^{(k)}(u|\delta^1) - \psi^{(m-k)}(u|\delta^2)\psi^{(k)}(u|\delta^2)|$$
$$\leqslant |\psi^{(m-k)}(u|\delta^1)|\cdot|\psi^{(k)}(u|\delta^1)-\psi^{(k)}(u|\delta^2)| + |\psi^{(k)}(u|\delta^2)|\cdot|\psi^{(m-k)}(u|\delta^1)-\psi^{(m-k)}(u|\delta^2)|$$
$$\leqslant (1+\sqrt{2}\pi^{m-k}||\delta^1||_{m-k})\cdot\sqrt{2}\pi^k||\delta^1-\delta^2||_k + (1+\sqrt{2}\pi^k||\delta^2||_k)\cdot\sqrt{2}\pi^{m-k}||\delta^1-\delta^2||_{m-k}$$
$$\leqslant \sqrt{2}[\pi^k + \pi^{m-k} + \sqrt{2}\pi^m(||\delta^1||_m + ||\delta^2||_m)]\cdot||\delta^1-\delta^2||_m.$$

Thus,

$$|\partial^m\psi^2(u|\delta^1)/\partial u^m - \partial^m\psi^2(u|\delta^2)/\partial u^m|$$
$$= \left|\sum_{k=0}^{m}\frac{m!}{k!(m-k)!}\left[\psi^{(m-k)}(u|\delta^1)\psi^{(k)}(u|\delta^1) - \psi^{(m-k)}(u|\delta^2)\psi^{(k)}(u|\delta^2)\right]\right|$$
$$\leqslant \sum_{k=0}^{m}\frac{m!}{k!(m-k)!}\sqrt{2}\left[\pi^k + \pi^{m-k} + \sqrt{2}\pi^m(||\delta^1||_m + ||\delta^2||_m)\right]\cdot||\delta^1-\delta^2||_m$$
$$\leqslant \left[2\sqrt{2}(1+\pi)^m + 2^{m+1}\pi^m(||\delta^1||_m + ||\delta^2||_m)\right]\cdot||\delta^1-\delta^2||_m.$$

Then,

$$|h^{(m)}(u|\delta^1) - h^{(m)}(u|\delta^2)| = (1-\epsilon_0)\left|\frac{\partial^m\psi^2(u|\delta^1)/\partial u^m}{1+\sum_{k=1}^\infty\delta_{1k}^2} - \frac{\partial^m\psi^2(u|\delta^2)/\partial u^m}{1+\sum_{k=1}^\infty\delta_{2k}^2}\right|$$
$$\leqslant (1-\epsilon_0)\left|\frac{\partial^m\psi^2(u|\delta^1)/\partial u^m}{1+\sum_{k=1}^\infty\delta_{1k}^2} - \frac{\partial^m\psi^2(u|\delta^1)/\partial u^m}{1+\sum_{k=1}^\infty\delta_{2k}^2}\right| + \frac{|\partial^m\psi^2(u|\delta^1)/\partial u^m - \partial^m\psi^2(u|\delta^2)/\partial u^m|}{1+\sum_{k=1}^\infty\delta_{2k}^2}$$
$$\leqslant \left|\sum_{k=1}^\infty(\delta_{2k}^2 - \delta_{1k}^2)\right|\cdot\left|h^{(m)}(u|\delta^1)\right| + \left[2\sqrt{2}(1+\pi)^m + 2^{m+1}\pi^m(||\delta^1||_m + ||\delta^2||_m)\right]\cdot||\delta^1-\delta^2||_m$$
$$\leqslant \left[2^m||\delta^1+\delta^2||_0(1+\sqrt{2}\pi^m||\delta^1||_m)^2 + 2\sqrt{2}(1+\pi)^m + 2^{m+1}\pi^m(||\delta^1||_m + ||\delta^2||_m)\right]\cdot||\delta^1-\delta^2||_m,$$

where the last inequality is from $\left|\sum_{k=1}^\infty(\delta_{2k}^2 - \delta_{1k}^2)\right| = |\sum_{k=1}^\infty(\delta_{2k}-\delta_{1k})(\delta_{2k}+\delta_{1k})| \leqslant ||\delta^1+\delta^2||_0\,\cdot$

$||\delta^1 - \delta^2||_0$ and the result (1).

(4) $\nabla_{\delta_j} h(u|\delta) = 2\left[(1-\epsilon_0)\psi(u|\delta)\sqrt{2}\cos j\pi u - \delta_j(h(u|\delta)-\epsilon_0)\right]/(1+\sum_{k=1}^{\infty}\delta_k^2)$.

$$
\begin{aligned}
&|\nabla_{\delta_j} h(u|\delta^1) - \nabla_{\delta_j} h(u|\delta^2)| \\
\leqslant &\left|\frac{2}{1+\sum_{k=1}^{\infty}\delta_{1k}^2} - \frac{2}{1+\sum_{k=1}^{\infty}\delta_{2k}^2}\right| \cdot \left|(1-\epsilon_0)\psi(u|\delta^1)\sqrt{2}\cos j\pi u - \delta_{1j}(h(u|\delta^1)-\epsilon_0)\right| + \\
&\frac{2}{1+\sum_{k=1}^{\infty}\delta_{2k}^2}\left|(1-\epsilon_0)\sqrt{2}\cos j\pi u\left[\psi(u|\delta^1)-\psi(u|\delta^2)\right] - \left[\delta_{1j}(h(u|\delta^1)-\epsilon_0) - \delta_{2j}(h(u|\delta^2)-\epsilon_0)\right]\right| \\
\leqslant &||\delta^1+\delta^2||_0\left[2\sqrt{2}(1+\sqrt{2}||\delta^1||_0) + (1+\sqrt{2}||\delta^1||_0)^2\right]\cdot||\delta^1-\delta^2||_0 + \\
&\frac{2}{1+\sum_{k=1}^{\infty}\delta_{2k}^2}\left[2||\delta^1-\delta^2||_0 + |\delta_{1j}-\delta_{2j}|\cdot|h(u|\delta^1)-\epsilon_0| + |\delta_{2j}|\cdot|h(u|\delta^1)-h(u|\delta^2)|\right] \\
\leqslant &\left\{(||\delta^1||_0+||\delta^2||_0)\left[2+2\sqrt{2}(1+\sqrt{2}||\delta^1||_0)+2(1+\sqrt{2}||\delta^1||_0)^2\right] + 4 + 2\sqrt{2} + 2(1+\sqrt{2}||\delta^1||_0)^2\right\}\cdot \\
&||\delta^1-\delta^2||_0 \leqslant [4(||\delta^1||_0+||\delta^2||_0+1)(\sqrt{2}+||\delta^1||_0)^2+1]\cdot||\delta^1-\delta^2||_0,
\end{aligned}
$$

where the first inequality is built on $1+\sum_{k=1}^{\infty}\delta_{1k}^2 \geqslant 1+\delta_{1j}^2 \geqslant 2|\delta_{1j}|$ and the last inequality holds because $2+2\sqrt{2}(1+\sqrt{2}||\delta^1||_0)+2(1+\sqrt{2}||\delta^1||_0)^2 \leqslant 4(\sqrt{2}+||\delta^1||_0)^2$ and $3+2\sqrt{2}+2(1+\sqrt{2}||\delta^1||_0)^2 < 4(\sqrt{2}+||\delta^1||_0)^2$.

When $m \geqslant 1$ is an integer, from Eq. (B.3),

$$
\begin{aligned}
&|\nabla_{\delta_j} h^{(m)}(u|\delta^1) - \nabla_{\delta_j} h^{(m)}(u|\delta^2)| \\
\leqslant & \frac{\left|2\sqrt{2}(1-\epsilon_0)\sum_{k=0}^m \frac{m!}{k!(m-k)!}(j\pi)^k \cos(j\pi u + \frac{k\pi}{2})[\psi^{(m-k)}(u|\delta^1) - \psi^{(m-k)}(u|\delta^2)]\right|}{1+\sum_{k=1}^\infty \delta_{1k}^2} + \\
& \left|2\sqrt{2}(1-\epsilon_0)\sum_{k=0}^m \frac{m!}{k!(m-k)!}(j\pi)^k \cos(j\pi u + \frac{k\pi}{2})\psi^{(m-k)}(u|\delta^2)\right| \cdot \left|\frac{1}{1+\sum_{k=1}^\infty \delta_{1k}^2} - \frac{1}{1+\sum_{k=1}^\infty \delta_{2k}^2}\right| \\
& + \frac{\left|2\delta_{1j}[h^{(m)}(u|\delta^1) - h^{(m)}(u|\delta^2)]\right|}{1+\sum_{k=1}^\infty \delta_{1k}^2} + \left|\frac{2\delta_{1j}}{1+\sum_{k=1}^\infty \delta_{1k}^2} - \frac{2\delta_{2j}}{1+\sum_{k=1}^\infty \delta_{2k}^2}\right| \cdot \left|h^{(m)}(u|\delta^2)\right| \\
\leqslant & 4\pi^m(1+j)^m||\delta^1-\delta^2||_m + [2^{1.5}j^m + 4(1+j)^m||\delta^2||_m]\pi^m||\delta^1+\delta^2||_0 \cdot ||\delta^1-\delta^2||_0 + \\
& \left[2^m||\delta^1+\delta^2||_0(1+\sqrt{2}\pi^m||\delta^2||_m)^2 + 2\sqrt{2}(1+\pi)^m + 2^{m+1}\pi^m(||\delta^1||_m + ||\delta^2||_m)\right] \cdot ||\delta^1-\delta^2||_m \\
& + (2+||\delta^1+\delta^2||_0)2^m(1+\sqrt{2}\pi^m||\delta^2||_m)^2 \cdot ||\delta^1-\delta^2||_0 \\
\leqslant & \pi^m \Big\{4(1+j)^m + [2^{1.5}j^m + 4(1+j)^m||\delta^2||_m + 2^{m+1}] \cdot (||\delta^1||_m + ||\delta^2||_m) + 2\sqrt{2}(1+\pi^{-1})^m \\
& + (2+2||\delta^1+\delta^2||_0)(\frac{2}{\pi})^m(1+\sqrt{2}\pi^m||\delta^2||_m)^2\Big\} \cdot ||\delta^1-\delta^2||_m,
\end{aligned}
$$

where the second inequality originates from the previous conclusions in this lemma and

$$
\begin{aligned}
&\left|\frac{\delta_{1i}}{1+\sum_{k=1}^\infty \delta_{1k}^2} - \frac{\delta_{2i}}{1+\sum_{k=1}^\infty \delta_{2k}^2}\right| \leqslant \frac{|\delta_{1i}-\delta_{2i}|}{1+\sum_{k=1}^\infty \delta_{1k}^2} + \left|\frac{\delta_{2i}}{1+\sum_{k=1}^\infty \delta_{1k}^2} - \frac{\delta_{2i}}{1+\sum_{k=1}^\infty \delta_{2k}^2}\right| \\
\leqslant & ||\delta^1-\delta^2||_0 + \frac{|\delta_{2i}| \cdot |\sum_{k=1}^\infty(\delta_{2k}^2 - \delta_{1k}^2)|}{(1+\sum_{k=1}^\infty \delta_{1k}^2)(1+\sum_{k=1}^\infty \delta_{2k}^2)} \\
\leqslant & ||\delta^1-\delta^2||_0 + \frac{|\delta_{2i}| \cdot ||\delta^1+\delta^2||_0 \cdot ||\delta^1-\delta^2||_0}{(1+\sum_{k=1}^\infty \delta_{1k}^2)(1+\sum_{k=1}^\infty \delta_{2k}^2)} \leqslant ||\delta^1-\delta^2||_0(1+\frac{1}{2}||\delta^1+\delta^2||_0).
\end{aligned}
\tag{B.4}
$$

(5) Differentiating Eq. (B.1) with respect to δ_i and arranging the order of the terms, we obtain

$$
\nabla_{\delta_i,\delta_j} h(u|\delta) = \frac{4(1-\epsilon_0)\cos i\pi u \cos j\pi u - 2[\delta_i \nabla_{\delta_j} h(u|\delta) + \delta_j \nabla_{\delta_i} h(u|\delta)] - 2[h(u|\delta)-\epsilon_0]1(i=j)}{1+\sum_{k=1}^\infty \delta_k^2}. \tag{B.5}
$$

Then the first result comes from

$$
\begin{aligned}
|\nabla_{\delta_i,\delta_j} h(u|\delta)| \leqslant & \frac{4 + 2\cdot 1(i=j)h(u|\delta) + 2\left|\delta_i \nabla_{\delta_j} h(u|\delta) + \delta_j \nabla_{\delta_i} h(u|\delta)\right|}{1+\sum_{k=1}^\infty \delta_k^2} \\
\leqslant & 4 + 2\cdot 1(i=j)(1+\sqrt{2}||\delta||_0)^2 + 8(1+\sqrt{2}||\delta||_0)^2.
\end{aligned}
$$

For $m \geqslant 1$, differentiate both sides of Eq. (B.2) with respect to δ_i:

$$2 \cdot 1(i=j)h^{(m)}(u|\delta) + 2\delta_j \nabla_{\delta_i} h^{(m)}(u|\delta) + 2\delta_i \nabla_{\delta_j} h^{(m)}(u|\delta) + (1+\sum_{k=1}^{\infty}\delta_k^2)\nabla_{\delta_i,\delta_j} h^{(m)}(u|\delta)$$
$$=4\pi^m(1-\epsilon_0)\sum_{k=0}^{m}\binom{m}{k} j^k i^{m-k}\cos(j\pi u + k\pi/2)\cos(i\pi u + (m-k)\pi/2).$$

As a result,

$$|\nabla_{\delta_i,\delta_j} h^{(m)}(u|\delta)| \leqslant \frac{4(i+j)^m\pi^m + 2|\delta_j\nabla_{\delta_i}h^{(m)}(u|\delta) + \delta_i\nabla_{\delta_j}h^{(m)}(u|\delta)| + 2\cdot 1(i=j)|h^{(m)}(u|\delta)|}{1+\sum_{k=1}^{\infty}\delta_k^2}$$
$$\leqslant 4(i+j)^m\pi^m + (1+\sqrt{2}\pi^m||\delta||_m)^2 2^m(2+4\pi^m j^m + 4\pi^m i^m)$$
$$\leqslant [1+(1+\sqrt{2}\pi^m||\delta||_m)^2]2^m(2+4\pi^m j^m + 4\pi^m i^m).$$

(6) By Eq. (B.4) and (B.5),

$$|\nabla_{\delta_i,\delta_j}h(u|\delta^1) - \nabla_{\delta_i,\delta_j}h(u|\delta^2)| \leqslant \left|\frac{1}{1+\sum_{k=1}^{\infty}\delta_{1k}^2} - \frac{1}{1+\sum_{k=1}^{\infty}\delta_{2k}^2}\right| \cdot$$
$$|4(1-\epsilon_0)\cos i\pi u\cos j\pi u - 2\cdot 1(i=j)[h(u|\delta^1)-\epsilon_0]| + \frac{2\cdot 1(i=j)|h(u|\delta^1)-h(u|\delta^2)|}{1+\sum_{k=1}^{\infty}\delta_{2k}^2} +$$
$$\left|\frac{2\delta_{1i}}{1+\sum_{k=1}^{\infty}\delta_{1k}^2} - \frac{2\delta_{2i}}{1+\sum_{k=1}^{\infty}\delta_{2k}^2}\right| \cdot |\nabla_{\delta_j}h(u|\delta^1)| + \frac{2|\delta_{2i}|}{1+\sum_{k=1}^{\infty}\delta_{2k}^2}\cdot|\nabla_{\delta_j}h(u|\delta^1) - \nabla_{\delta_j}h(u|\delta^2)| +$$
$$\left|\frac{2\delta_{1j}}{1+\sum_{k=1}^{\infty}\delta_{1k}^2} - \frac{2\delta_{2j}}{1+\sum_{k=1}^{\infty}\delta_{2k}^2}\right| \cdot |\nabla_{\delta_i}h(u|\delta^1)| + \frac{2|\delta_{2j}|}{1+\sum_{k=1}^{\infty}\delta_{2k}^2}\cdot|\nabla_{\delta_i}h(u|\delta^1) - \nabla_{\delta_i}h(u|\delta^2)|$$
$$\leqslant ||\delta^1-\delta^2||_0 \cdot ||\delta^1+\delta^2||_0[4+2(1+\sqrt{2}||\delta^1||_0)^2] + 2\left[||\delta^1+\delta^2||_0(1+\sqrt{2}||\delta^1||_0)^2 + 2\sqrt{2} +\right.$$
$$\left.2(||\delta^1|| + ||\delta^2||_0)\right] \cdot ||\delta^1-\delta^2||_0 + 2\cdot||\delta^1-\delta^2||_0(2+||\delta^1+\delta^2||_0)\cdot 4(1+\sqrt{2}||\delta^1||_0)^2$$
$$+ 2||\delta^1-\delta^2||_0\cdot[(4||\delta^1||_0 + 4||\delta^2||_0 + 4)(\sqrt{2}+||\delta^1||_0)^2 + 1]$$
$$\leqslant ||\delta^1-\delta^2||_0\cdot(2+||\delta^1||_0+||\delta^2||_0)\left[12+8\sqrt{2}||\delta^1||_0 + 16(1+\sqrt{2}||\delta^1||_0)^2\right].$$

(7)

$$\left|\frac{h'(u|\delta)}{h(u|\delta)}\right| = \left|\frac{-(1-\epsilon_0)(1+\sum_{k=1}^{\infty}\delta_k^2)^{-1}2(1+\sum_{k=1}^{\infty}\delta_k\sqrt{2}\cos k\pi u)\sum_{k=1}^{\infty}\delta_k\sqrt{2}k\pi\sin k\pi u}{(1-\epsilon_0)(1+\sum_{k=1}^{\infty}\delta_k^2)^{-1}(1+\sum_{k=1}^{\infty}\delta_k\sqrt{2}\cos k\pi u)^2+\epsilon_0}\right|$$
$$\leqslant \left|\frac{2(1-\epsilon_0)(1+\sum_{k=1}^{\infty}\delta_k^2)^{-1}(1+\sum_{k=1}^{\infty}\delta_k\sqrt{2}\cos k\pi u)\sum_{k=1}^{\infty}\delta_k\sqrt{2}k\pi\sin k\pi u}{2\sqrt{\epsilon_0(1-\epsilon_0)}(1+\sum_{k=1}^{\infty}\delta_k^2)^{-1/2}(1+\sum_{k=1}^{\infty}\delta_k\sqrt{2}\cos k\pi u)}\right| \tag{B.6}$$
$$\leqslant \epsilon_0^{-1/2}\left|\sum_{k=1}^{\infty}\delta_k\sqrt{2}k\pi\sin k\pi u\right| \leqslant \pi||\delta||_1(2/\epsilon_0)^{1/2}.$$

(8)By conclusion (3) in this lemma,

$$\left|\frac{1}{h(u|\delta^1)}-\frac{1}{h(u|\delta^2)}\right| = \frac{|h(u|\delta^1)-h(u|\delta^2)|}{h(u|\delta^1)h(u|\delta^2)}$$
$$\leqslant \epsilon_0^{-2}[||\delta^1+\delta^2||_0(1+\sqrt{2}||\delta^1||_0)^2+2\sqrt{2}+2(||\delta^1||_0+||\delta^2||_0)]||\delta^1-\delta^2||_0.$$

□

Proof of Lemma B.2: (1) & (2) The results are derived by Lagrange's mean value theorem.

(3) The result is obtained by differentiating the expression of $H(u|\delta)$ with respect to δ_k.

(4) With Lemma B.1 (2), Lebesgue's dominated convergence theorem is applicable:

$$\sup_u \left|\frac{\nabla_{\delta_k}H(u|\delta)}{u}\right| = \sup_u \frac{1}{u}\left|\nabla_{\delta_k}\int_0^u h(u|\delta)du\right| = \sup_u \frac{1}{u}\left|\int_0^u \nabla_{\delta_k}h(u|\delta)du\right| \leqslant 4(1+\sqrt{2}||\delta||_0)^2.$$

(5) $|\nabla_{\delta_k}H(u|\delta)/H(u|\delta)| \leqslant 1+\frac{2u/H(u|\delta)}{1+\sum_{j=1}^{\infty}\delta_j^2}|\delta_k+\sqrt{2}\frac{\sin k\pi u}{k\pi u}+\sum_{j=1}^{\infty}\delta_j\frac{\sin(k+j)\pi u}{(k+j)\pi u}+\sum_{j\neq k}\delta_j\frac{\sin(k-j)\pi u}{(k-j)\pi u}| \leqslant 1+\epsilon_0^{-1}(2\sqrt{2}+4||\delta||_0)$.

(6) Again, with Lemma B.1 (2), Lebesgue's dominated convergence theorem is applicable:

$$\sup_{0<u<1}\frac{|\nabla_{\delta_k}H(u|\delta^1)-\nabla_{\delta_k}H(u|\delta^2)|}{u} = \sup_{0<u<1}\frac{1}{u}\left|\int_0^u[\nabla_{\delta_k}h(v|\delta^1)-\nabla_{\delta_k}h(v|\delta^2)]dv\right|$$
$$\leqslant \sup_{0<v<1}\left|\nabla_{\delta_k}h(v|\delta^1)-\nabla_{\delta_k}h(v|\delta^2)\right| \leqslant \left[4(||\delta^1||_0+||\delta^2||_0+1)(\sqrt{2}+||\delta^1||_0)^2+1\right]\cdot||\delta^1-\delta^2||_0.$$

(7) With Lemma B.1 (5), Lebesgue's dominated convergence theorem is applicable. Then conclusion is a result of Lemma B.1 (5).

(8) With Lemma B.1 (5), Lebesgue's dominated convergence theorem is applicable. Then conclusion is a result of Lemma B.1 (6).

(9) By Lemma B.1 (2) and conclusion (4) in this lemma, the conclusion holds because $\frac{H(u|\delta)}{u} \geqslant \epsilon_0$

and $u\partial[\nabla_{\delta_k} H(u|\delta)/u]/\partial u = \nabla_{\delta_k} h(u|\delta) - \nabla_{\delta_k} H(u|\delta)/u$.

(10) Because $\sup_u |h(u|\delta)| \leqslant (1+\sqrt{2}||\delta||_0)^2$, the conclusion is implied by

$$u\frac{\partial}{\partial u}\frac{\nabla_{\delta_k} H(u|\delta)}{H(u|\delta)} = \frac{\nabla_{\delta_k} h(u|\delta)}{H(u|\delta)/u} - \frac{\nabla_{\delta_k} H(u|\delta)}{u}\frac{h(u|\delta)}{[H(u|\delta)/u]^2}.$$

(11) Because $\frac{H(u|\delta)}{u} \geqslant \epsilon_0$, $\sup_u |\nabla_{\delta_k,\delta_j} h(u|\delta)| \leqslant 4 + 10(1+\sqrt{2}||\delta||_0)^2$, $\sup_u |\nabla_{\delta_k,\delta_j} H(u|\delta)/u| \leqslant 4 + 10(1+\sqrt{2}||\delta||_0)^2$, and $\sup_u |h(u|\delta)| \leqslant (1+\sqrt{2}||\delta||_0)^2$, the conclusion is deduced from

$$\begin{aligned} u\frac{\partial}{\partial u}\frac{\nabla_{\delta_k,\delta_j} H(u|\delta)}{H(u|\delta)} &= \frac{\nabla_{\delta_k,\delta_j} h(u|\delta)}{H(u|\delta)/u} - u\frac{\nabla_{\delta_k,\delta_j} H(u|\delta)\cdot h(u|\delta)}{H(u|\delta)^2} \\ =&\nabla_{\delta_k,\delta_j} h(u|\delta)\cdot\frac{1}{H(u|\delta)/u} - \frac{\nabla_{\delta_k,\delta_j} H(u|\delta)}{u}\frac{h(u|\delta)}{[H(u|\delta)/u]^2}. \end{aligned}$$

□

Proof of Lemma B.3: (1) $\psi_1'(u|\delta) = 1/H(u|\delta) - uh(u|\delta)/H(u|\delta)^2$. Because $\frac{H(u|\delta)}{u} \geqslant \epsilon_0$ and $\sup_v |h(v|\delta)| \leqslant (1+\sqrt{2}||\delta||_0)^2$, $\sup_u |\psi_1'(u|\delta)u| \leqslant \frac{1}{H(u|\delta)/u} + \frac{h(u|\delta)}{[H(u|\delta)/u]^2} \leqslant C(1+\sqrt{2}||\delta||_0)^2$ for some $C > 0$.

(2) $\nabla_{\delta_j}\psi_1(u|\delta) = -u\nabla_{\delta_j} H(u|\delta)/H(u|\delta)^2$. Because $\frac{H(u|\delta)}{u} \geqslant \epsilon_0$ and $\sup_v |\nabla_{\delta_j} h(v|\delta)| \leqslant 4(1+\sqrt{2}||\delta||_0)^2$,

$$|\nabla_{\delta_j}\psi_1(u|\delta)| \leqslant \frac{u\cdot u\sup_{0<v<1}|\nabla_{\delta_j} h(v|\delta)|}{H(u|\delta)^2} \leqslant \frac{\sup_{0<v<1}|\nabla_{\delta_j} h(v|\delta)|}{[H(u|\delta)/u]^2} \leqslant 4\epsilon_0^{-2}(1+\sqrt{2}||\delta||_0)^2.$$

(3) By Lemma B.1 (3),

$$\begin{aligned} |\psi_1(u|\delta^1) - \psi_1(u|\delta^2)| &= \frac{u|H(u|\delta^2) - H(u|\delta^1)|}{H(u|\delta^1)H(u|\delta^2)} \leqslant \frac{\sup_{0\leqslant v\leqslant 1}|h(v|\delta^2) - h(v|\delta^1)|}{[H(u|\delta^1)/u][H(u|\delta^2)/u]} \\ \leqslant&\epsilon_0^{-2}\left[(||\delta^1||_0 + ||\delta^2||_0)\left((1+\sqrt{2}||\delta^1||_0)^2 + 2\right) + 2\sqrt{2}\right]\cdot||\delta^1 - \delta^2||_0. \end{aligned}$$

(4) $\psi_1'(u|\delta) = \frac{1}{H(u|\delta)} - \frac{uh(u|\delta)}{H(u|\delta)^2}$. Because $\sup_{0\leqslant v\leqslant 1}|h(v|\delta^2) - h(v|\delta^1)| \leqslant [(||\delta^1||_0 + ||\delta^2||_0)((1+$

$\sqrt{2}||\delta^1||_0)^2+2)+2\sqrt{2}]\cdot||\delta^2-\delta^1||_0$ and $\sup_v|h(v|\delta)|\leqslant(1+\sqrt{2}||\delta||_0)^2$,

$$\begin{aligned}
&|\psi_1'(u|\delta^1)-\psi_1'(u|\delta^2)|u\\
\leqslant&\frac{u\cdot u\sup_{0\leqslant v\leqslant 1}|h(v|\delta^2)-h(v|\delta^1)|}{H(u|\delta^1)H(u|\delta^2)}+\frac{u^2|h(u|\delta^1)H(u|\delta^2)^2-h(u|\delta^2)H(u|\delta^1)^2|}{H(u|\delta^1)^2H(u|\delta^2)^2}\\
\leqslant&\epsilon_0^{-2}\sup_{0\leqslant v\leqslant 1}|h(v|\delta^2)-h(v|\delta^1)|+\frac{u^2|h(u|\delta^1)-h(u|\delta^2)|}{H(u|\delta^1)^2}+\frac{u^2h(u|\delta^2)\cdot|H(u|\delta^2)^2-H(u|\delta^1)^2|}{H(u|\delta^1)^2H(u|\delta^2)^2}\\
\leqslant&2\epsilon_0^{-2}\sup_{0\leqslant v\leqslant 1}|h(v|\delta^2)-h(v|\delta^1)|+(1+\sqrt{2}||\delta||_0)^2\epsilon_0^{-2}\frac{[H(u|\delta^2)+H(u|\delta^1)]\cdot|H(u|\delta^2)-H(u|\delta^1)|}{H(u|\delta^1)H(u|\delta^2)}\\
\leqslant&\epsilon_0^{-2}\sup_{0\leqslant v\leqslant 1}|h(v|\delta^2)-h(v|\delta^1)|\left\{2+(1+\sqrt{2}||\delta||_0)^2[\frac{1}{H(u|\delta^1)}+\frac{1}{H(u|\delta^2)}]u\right\}\\
\leqslant&2\epsilon_0^{-3}[\epsilon_0+(1+\sqrt{2}||\delta||_0)^2]\left[(||\delta^1||_0+||\delta^2||_0)((1+\sqrt{2}||\delta^1||_0)^2+2)+2\sqrt{2}\right]\cdot||\delta^2-\delta^1||_0.
\end{aligned}$$

(5) Because $\frac{H(u|\delta)}{u}\geqslant\epsilon_0$, $\sup_v|h(v|\delta^2)-h(v|\delta^1)|\leqslant\left[(||\delta^1||_0+||\delta^2||_0)((1+\sqrt{2}||\delta^1||_0)^2+2)+2\sqrt{2}\right]\cdot||\delta^2-\delta^1||_0$, $\sup_v|\nabla_{\delta_j}H(v|\delta)/v|\leqslant 4(1+\sqrt{2}||\delta||_0)^2$, and $\sup_v|\nabla_{\delta_j}H(v|\delta^1)-\nabla_{\delta_j}H(v|\delta^2)|/v\leqslant[4(||\delta^1||_0+||\delta^2||_0+1)(\sqrt{2}+||\delta^1||_0)^2+1]\cdot||\delta^1-\delta^2||_0$, we have

$$\begin{aligned}
&|\nabla_{\delta_j}\psi_1(u|\delta^1)-\nabla_{\delta_j}\psi_1(u|\delta^2)|\\
\leqslant&\frac{u|\nabla_{\delta_j}H(u|\delta^1)-\nabla_{\delta_j}H(u|\delta^2)|}{H(u|\delta^1)^2}+\frac{u|\nabla_{\delta_j}H(u|\delta^2)|\cdot|[H(u|\delta^1)+H(u|\delta^2)][H(u|\delta^1)-H(u|\delta^2)]|}{H(u|\delta^1)^2H(u|\delta^2)^2}\\
\leqslant&\epsilon_0^{-2}\left[4(||\delta^1||_0+||\delta^2||_0+1)(\sqrt{2}+||\delta^1||_0)^2+1\right]\cdot||\delta^1-\delta^2||_0+\epsilon_0^{-2}4(1+\sqrt{2}||\delta^2||_0)^2\cdot\\
&\left\{\left[\frac{u}{H(u|\delta^1)}+\frac{u}{H(u|\delta^2)}\right]\left[(||\delta^1||_0+||\delta^2||_0)((1+\sqrt{2}||\delta^1||_0)^2+2)+2\sqrt{2}\right]\right\}\cdot||\delta^2-\delta^1||_0\\
\leqslant&C(1+||\delta^2||_0)^2(||\delta^1||_0+||\delta^2||_0+1)(1+||\delta^1||_0)^2\cdot||\delta^2-\delta^1||_0
\end{aligned}$$

for some constant $C>0$. □

C. Identification and Consistency–Proofs for Sections 2 & 3

C.1. The Proof for Section 2

Proof of Lemma 1: We will show that if $L_N(\theta)=L_N(\theta_0)$ a.s., then θ=θ_0. Since $\epsilon_{i,N}$'s are *i.i.d.* and have support $\mathbb{R}^N$, $P(y_{1,N}=y_{2,N}=\cdots=y_{N,N}=0)>0$ and $P(y_{1,N}>0,y_{2,N}>0,\cdots y_{N,N}>0)>0$. We will discuss the identification by $K^0=1$ and $K^0>1$.

(1) When $K^0=1$, for the event that $y_{1,N}=y_{2,N}=\cdots=y_{N,N}=0$, $L_N(\theta)=L_N(\theta_0)$ a.s.

implies $\prod_{i=1}^N F_0(-x_{i,N}\beta_0) = \prod_{i=1}^N F(-x_{i,N}\beta)$ a.s.. Since support$(x_{1,N}, x_{2,N}, \cdots, x_{N,N}) = \mathbb{R}^N$, we can consider the case $x_{i,N} \to x_{1,N}$ for all $2 \leqslant i \leqslant N$: $\prod_{i=1}^N F_0(-x_{1,N}\beta_0) = \prod_{i=1}^N F(-x_{1,N}\beta)$ a.s.. That is to say, $F_0(-x_{1,N}\beta_0) = F(-x_{1,N}\beta)$ a.s.. If $\beta = 0$ but $\beta_0 \neq 0$ or vice versa, the equality of the probabilities is impossible. So, consider $\beta \neq 0$ and $\beta_0 \neq 0$. Thus, $F(t) = F_0(t\frac{\beta_0}{\beta})$ and, consequently, $f(t) = \frac{\beta_0}{\beta} f_0(t\frac{\beta_0}{\beta})$.

On the other hand, for the event $y_{1,N} > 0$, $y_{2,N} > 0, \cdots, y_{N,N} > 0$, $L_N(\theta) = L_N(\theta_0)$ a.s. implies $|I_N - \lambda_0 W_N| \prod_{i=1}^N f_0(y_{i,N} - \lambda_0 w_{i\cdot,N} Y_N - x_{i,N}\beta_0) = |I_N - \lambda W_N| \prod_{i=1}^N f(y_{i,N} - \lambda w_{i\cdot,N} Y_N - x_{i,N}\beta)$. By $f(t) = \frac{\beta_0}{\beta} f_0(t\frac{\beta_0}{\beta})$,

$$\begin{aligned} &|I_N - \lambda_0 W_N| \prod_{i=1}^N f_0(y_{i,N} - \lambda_0 w_{i\cdot,N} Y_N - x_{i,N}\beta_0) \\ =&|I_N - \lambda W_N| (\frac{\beta_0}{\beta})^N \prod_{i=1}^N f_0(\frac{\beta_0}{\beta} y_{i,N} - \frac{\beta_0 \lambda}{\beta} w_{i\cdot,N} Y_N - x_{i,N}\beta_0). \end{aligned} \tag{C.1}$$

Letting all $y_{i,N} \downarrow 0$, we obtain $|I_N - \lambda_0 W_N| \prod_{i=1}^N f_0(-x_{i,N}\beta_0) = |I_N - \lambda W_N| (\frac{\beta_0}{\beta})^N \prod_{i=1}^N f_0(-x_{i,N}\beta_0)$. It follows from $f_0(x) > 0$ that $|I_N - \lambda_0 W_N| = |I_N - \lambda W_N| (\frac{\beta_0}{\beta})^N$. Assumption 3 implies that $|I_N - \lambda_0 W_N| \neq 0$. Thus, Eq. (C.1) becomes $\prod_{i=1}^N f_0(y_{i,N} - \lambda_0 w_{i\cdot,N} Y_N - x_{i,N}\beta_0) = \prod_{i=1}^N f_0(\frac{\beta_0}{\beta} y_{i,N} - \frac{\beta_0 \lambda}{\beta} w_{i\cdot,N} Y_N - x_{i,N}\beta_0)$ a.s.. Take logarithm and differentiate both sides with respect to $x_{i,N}$ and denote $u(x) \equiv d \ln f_0(x)/dx$, then

$$u(y_{i,N} - \lambda_0 w_{i\cdot,N} Y_N - x_{i,N}\beta_0) = u(\frac{\beta_0}{\beta} y_{i,N} - \frac{\beta_0 \lambda}{\beta} w_{i\cdot,N} Y_N - x_{i,N}\beta_0) \tag{C.2}$$

a.s. as $\beta \neq 0$. Letting $y_{j,N} \downarrow 0$ for all $j \neq i$, we have

$$u(y_{i,N} - x_{i,N}\beta_0) = u(\frac{\beta_0}{\beta} y_{i,N} - x_{i,N}\beta_0) = u(y_{i,N} - x_{i,N}\beta_0 + (\frac{\beta_0}{\beta} - 1) y_{i,N}) \tag{C.3}$$

a.s.. As $f_0(x)$ is a density function, $u(x)$ is not a constant function. There are intervals in the domains of $u(\cdot)$ such that $u(\cdot)$ is strictly monotonic. As support$(x_{i,N}\beta_0) = \mathbb{R}$ and support$(y_{i,N}) = (0, \infty)$, we must have $\frac{\beta_0}{\beta} - 1 = 0$ from Eq. (C.3). Thus, Eq. (C.2) becomes $u(y_{i,N} - \lambda_0 w_{i\cdot,N} Y_N - x_{i,N}\beta_0) = u(y_{i,N} - \lambda_0 w_{i\cdot,N} Y_N - x_{i,N}\beta_0 + (\lambda_0 - \lambda) w_{i\cdot,N} Y_N)$. Since $W_N \neq 0$, there must be an i such that $w_{i\cdot,N} Y_N$ has values in $[0, \infty)$. Thus, the strict monotonicity of $u(\cdot)$ in some intervals implies that $\lambda_0 = \lambda$. Finally, $F(t) = F_0(t\frac{\beta_0}{\beta})$ for all t and $\beta = \beta_0$ imply $F(\cdot) \equiv F_0(\cdot)$.

(2) For $K^0 > 1$, similarly, as in a preceding argument, $F_0(-x_{i,N}\beta_0) = F(-x_{i,N}\beta)$ a.s. on $x_{i,N}$, i.e., $F_0(-x_{i1,N}\beta_{01} - x_{i,\sim,N}\beta_{0,\sim}) = F(-x_{i1,N}\beta_1 - x_{i,\sim,N}\beta_\sim)$ a.s.. Without loss of generality, we may consider the case that neither β_{01} nor β_1 is zero. By fixing $x_{i,\sim,N}$, because support$(x_{i1,N}|x_{i,\sim,N}) = \mathbb{R}$ a.s., it implies $F_0(t\frac{\beta_{01}}{\beta_1} - x_{i,\sim,N}\beta_{0,\sim}) = F(t - x_{i,\sim,N}\beta_\sim)$ for all t. Differentiate this equation with respect to t, $\frac{\beta_{01}}{\beta_1} f_0(t\frac{\beta_{01}}{\beta_1} - x_{i,\sim,N}\beta_{0,\sim}) = f(t - x_{i,\sim,N}\beta_\sim)$. Plugging this relationship into the likelihood function with all $y_{i,N}$'s being positive, we obtain

$$\begin{aligned} &|I_N - \lambda_0 W_N| \prod_{i=1}^{N} f_0(y_{i,N} - \lambda_0 w_{i\cdot,N} Y_N - x_{i1,N}\beta_{01} - x_{i,\sim,N}\beta_{0,\sim}) \\ =&|I_N - \lambda W_N| (\frac{\beta_{01}}{\beta_1})^N \prod_{i=1}^{N} f_0(\frac{\beta_{01}}{\beta_1} y_{i,N} - \frac{\lambda\beta_{01}}{\beta_1} w_{i\cdot,N} Y_N - x_{i1,N}\beta_{01} - x_{i,\sim,N}\beta_{0,\sim}). \end{aligned} \tag{C.4}$$

Letting all $y_{i,N} \downarrow 0$, we have $|I_N - \lambda_0 W_N| = |I_N - \lambda W_N|(\frac{\beta_{01}}{\beta_1})^N$. Thus, in turn, Eq. (C.4) implies $\prod_{i=1}^N f_0(y_{i,N} - \lambda_0 w_{i\cdot,N} Y_N - x_{i,N}\beta_0) = \prod_{i=1}^N f_0(\frac{\beta_{01}}{\beta_1} y_{i,N} - \frac{\lambda\beta_{01}}{\beta_1} w_{i\cdot,N} Y_N - x_{i,N}\beta_0)$. Taking logarithm and differentiating this equation with respect to $x_{i,1,N}$ on both sides of the equation, we obtain $u(y_{i,N} - \lambda_0 w_{i\cdot,N} Y_N - x_{i,N}\beta_0) = u(\frac{\beta_{01}}{\beta_1} y_{i,N} - \frac{\lambda\beta_{01}}{\beta_1} w_{i\cdot,N} Y_N - x_{i,N}\beta_0)$. Let $y_{j,N} \downarrow 0$ for all $j \neq i$, $u(y_{i,N} - x_{i,N}\beta_0) = u(\frac{\beta_{01}}{\beta_1} y_{i,N} - x_{i,N}\beta_0)$, and therefore, $\beta_{01} = \beta_1$. It follows that $u(y_{i,N} - \lambda_0 w_{i\cdot,N} Y_N - x_{i,N}\beta_0) = u(y_{i,N} - \lambda_0 w_{i\cdot,N} Y_N - x_{i,N}\beta_0 + (\lambda_0 - \lambda) w_{i\cdot,N} Y_N)$. As $W_N \neq 0$, there must be an i such that $w_{i\cdot,N} Y_N$ has support $[0, \infty)$. As $u(\cdot)$ is strictly monotonic on some intervals, $\lambda_0 = \lambda$.

With $\beta_{01} = \beta_1$, $F_0(-x_{i1,N}\beta_{01} - x_{i,\sim,N}\beta_{0,\sim}) = F(-x_{i1,N}\beta_{01} - x_{i,\sim,N}\beta_\sim)$ a.s.. Let $\bar{t}_i = -x_{i1,N}\beta_{01} - x_{i,\sim,N}\beta_{0,\sim}$, which has support $\mathbb{R}$. Hence, $F_0(\bar{t}_i) = F(\bar{t}_i + x_{i,\sim,N}(\beta_{0,\sim} - \beta_\sim))$ a.s. for all $\bar{t}_i$. Notice that conditional on $\bar{t}_i$, the left hand side is nonstochastic, and therefore, by $u(\cdot)$ being strictly monotonic on some intervals, the right hand side must also be nonstochastic, which is possible only if $0 = \mathrm{var}(x_{i,\sim,N}(\beta_{0,\sim} - \beta_\sim)) = (\beta_{0,\sim} - \beta_\sim)' \mathrm{var}(x_{i,\sim,N})(\beta_{0,\sim} - \beta_\sim)$. Because $\mathrm{var}(x_{i,\sim,N})$ has full rank by Assumption 4 (2.2), $\beta_{0,\sim} = \beta_\sim$. Finally, it follows from $F_0(t) = F(t - x_{i,\sim,N}(\beta_{0,\sim} - \beta_\sim))$ that $F_0(t) = F(t)$ for all $t \in \mathbb{R}$, i.e., the CDF is also identifiable. □

Proof of Lemma 2: (1) With $K^0 = 2$, similarly to the proof of Lemma 1, $F_0(-x_{i,N}\beta_0) = F(-x_{i,N}\beta)$, i.e., $F_0(-\beta_{01} - x_{i2,N}\beta_{02}) = F(-\beta_1 - x_{i2,N}\beta_2)$. Let $\overline{F}_0(x) \equiv F_0(-\beta_{01} - x)$ and $\overline{F}(x) \equiv F(-\beta_1 - x)$. By the same argument as that in the proof for Lemma 1 with $\bar{F}_0$ in place of F_0, $\beta_{02} = \beta_2$, $\lambda_0 = \lambda$ and $\overline{F}_0(x) \equiv \overline{F}(x)$. Notice that $0.5 = F_0(0) = F_0(-\beta_{01} + \beta_{01}) = \overline{F}_0(\beta_{01}) =$

$\overline{F}(\beta_{01}) = F(-\beta_1 + \beta_{01})$, and $F(\cdot)$ is strictly increasing because F_0 is, $F^{-1}(0.5) = 0 = -\beta_1 + \beta_{01}$, i.e., $\beta_1 = \beta_{01}$.

(2) For $K^0 > 2$, similarly, $F_0(-\beta_{01} - x_{i2,N}\beta_{02} - x_{i,\sim}\beta_{0,\sim}) = F(-\beta_1 - x_{i2,N}\beta_2 - x_{i,\sim}\beta_{\sim})$. Let $\overline{F}_0(x) \equiv F_0(-\beta_{01} - x)$ and $\overline{F}(x) \equiv F(-\beta_1 - x)$. The rest of the proof combines those in part (2) proof for Lemma 1 and part (1) proof of this lemma. □

Proof of Lemma 3: In this proof, for any matrix $A = (a_{ij})$, denote $|A| \equiv (|a_{ij}|)$, consisting of the absolute value of each entry. Let $(m_{ij,N}) \equiv (I_N - |\lambda_0 W_N|)^{-1}$. Notice $||(I_N - |\lambda_0 W_N|)^{-1}||_\infty = ||\sum_{k=0}^{\infty} |\lambda_0 W_N|^k||_\infty \leqslant \sum_{k=0}^{\infty} ||\lambda_0 W_N||_\infty^k \leqslant (1 - \sup_N ||\lambda_0 W_N||_\infty)^{-1}$. Because $|y_{i,N}| \leqslant \sum_{j=1}^N m_{ij,N}|x_{j,N}\beta_0 + \epsilon_{j,N}|$ from Proposition 1 in XL (2015b),

$$
\begin{aligned}
&\mathrm{E}\, e^{\gamma|y_{i,N}|} \leqslant \mathrm{E}\exp(\gamma\sum_{j=1}^N m_{ij,N}|x_{j,N}\beta_0 + \epsilon_{j,N}|) \\
=& \mathrm{E}\prod_{j=1}^N [\exp(\gamma|x_{j,N}\beta_0 + \epsilon_{j,N}|\sum_{k=1}^N m_{ik,N})]^{m_{ij,N}/\sum_{k=1}^N m_{ik,N}} \\
\leqslant& \mathrm{E}\sum_{j=1}^N \frac{m_{ij,N}}{\sum_{k=1}^N m_{ik,N}} \exp(\gamma|x_{j,N}\beta_0 + \epsilon_{j,N}|\sum_{k=1}^N m_{ik,N}) \\
\leqslant& \sum_{j=1}^N \frac{m_{ij,N}}{\sum_{k=1}^N m_{ik,N}} \mathrm{E}\exp[\gamma|x_{j,N}\beta_0 + \epsilon_{j,N}|/(1 - \sup_N ||\lambda_0 W_N||_\infty)] \\
\leqslant& \sum_{j=1}^N \frac{m_{ij,N}}{\sum_{k=1}^N m_{ik,N}} \mathrm{E}\exp[\frac{\gamma|x_{j,N}\beta_0|}{1 - \sup_N ||\lambda_0 W_N||_\infty}] \cdot \mathrm{E}\exp[\frac{\gamma|\epsilon_{j,N}|}{1 - \sup_N ||\lambda_0 W_N||_\infty}] \\
\leqslant& \sup_{j,N} \mathrm{E}\exp[\frac{\gamma|x_{j,N}\beta_0|}{1 - \sup_N ||\lambda_0 W_N||_\infty}] \cdot \sup_{j,N} \mathrm{E}\exp[\frac{\gamma|\epsilon_{j,N}|}{1 - \sup_N ||\lambda_0 W_N||_\infty}] < \infty.
\end{aligned}
$$

where the second inequality comes from the general inequality of arithmetic and geometric means (Steele, 2004, p. 23). Similarly,

$$
\mathrm{E}[e^{\gamma|y_{i,N}|}|X_N] \leqslant \exp(\gamma\sum_{j=1}^N m_{ij,N}|x_{j,N}\beta_0|) \sup_{j,N} \mathrm{E}\exp[\frac{\gamma|\epsilon_{j,N}|}{1 - \sup_N ||\lambda_0 W_N||_\infty}]. \tag{C.5}
$$

Next, consider $z_{i,N}(\lambda, \beta) \equiv y_{i,N} - \lambda w_{i\cdot,N} Y_N - x_{i,N}\beta$. Let $\delta_{ij} \equiv 1(i = j)$ be a Kronecker delta. When $0 < \gamma \leqslant (1 - \sup_N ||\lambda_0 W_N||_\infty)(1 + \zeta)^{-1}\min(\frac{1}{2}\gamma_x, \gamma_\epsilon)$, $C_1 \equiv \sup_{\beta,i,N} \mathrm{E}^{1/2}(e^{2\gamma|x_{i,N}\beta|}) < \infty$, $C_2 \equiv \mathrm{E}\exp[\frac{\gamma(1+\zeta)|\epsilon_{j,N}|}{1-\sup_N ||\lambda_0 W_N||_\infty}] < \infty$, and

$$
\begin{aligned}
&\sup_{\lambda,\beta,i,N} \mathrm{E}\exp(\gamma|z_{i,N}(\lambda,\beta)|) \leqslant \sup_{\lambda,\beta,i,N} \mathrm{E}\{\, e^{\gamma|x_{i,N}\beta|}\, \mathrm{E}[e^{\gamma|y_{i,N}-\lambda w_{i\cdot,N}Y_N|} \Big| X_N]\} \\
&\leqslant \sup_{\beta,i,N} \mathrm{E}^{1/2}[e^{2\gamma|x_{i,N}\beta|}] \sup_{\lambda,i,N} \mathrm{E}^{1/2}\left\{\mathrm{E}^2\left[\exp(\gamma\sum_{j=1}^{N}|\delta_{ij}-\lambda w_{ij,N}|\cdot|y_{j,N}|)\Big| X_N\right]\right\} \\
&= C_1 \sup_{\lambda\in\Lambda,i,N} \mathrm{E}^{1/2}\left\{\mathrm{E}^2\left[\exp\left(\gamma\sum_{j=1}^{N}\frac{|\delta_{ij}-\lambda w_{ij,N}|}{\sum_{k=1}^{N}|\delta_{ik}-\lambda w_{ik,N}|}|y_{j,N}|\sum_{k=1}^{N}|\delta_{ik}-\lambda w_{ik,N}|\right)\Big| X_N\right]\right\} \\
&\leqslant C_1 \sup_{\lambda\in\Lambda,i,N} \mathrm{E}^{1/2}\left\{\left[\sum_{j=1}^{N}\frac{|\delta_{ij}-\lambda w_{ij,N}|}{\sum_{k=1}^{N}|\delta_{ik}-\lambda w_{ik,N}|}\,\mathrm{E}\left[\exp(\gamma|y_{j,N}|\sum_{k=1}^{N}|\delta_{ik}-\lambda w_{ik,N}|)\Big| X_N\right]\right]^2\right\} \\
&\leqslant C_1C_2 \sup_{\lambda\in\Lambda,i,N} \mathrm{E}^{1/2}\left\{\left[\sum_{j=1}^{N}\frac{|\delta_{ij}-\lambda w_{ij,N}|}{\sum_{k=1}^{N}|\delta_{ik}-\lambda w_{ik,N}|}\exp\left(\gamma(1+\zeta)\sum_{k=1}^{N}m_{jk,N}|x_{k,N}\beta_0|\right)\right]^2\right\} \\
&\leqslant C_1C_2 \sup_{\lambda\in\Lambda,i,N}\sum_{j=1}^{N}\frac{|\delta_{ij}-\lambda w_{ij,N}|}{\sum_{k=1}^{N}|\delta_{ik}-\lambda w_{ik,N}|}\mathrm{E}^{1/2}\left\{\exp\left[2\gamma(1+\zeta)\sum_{k=1}^{N}m_{jk,N}|x_{k,N}\beta_0|\right]\right\} \\
&= C_1C_2 \sup_{i,N}\mathrm{E}^{1/2}\left\{\exp\left[2\gamma(1+\zeta)(\sum_{j=1}^{N}m_{ij,N})\sum_{k=1}^{N}\frac{m_{ik,N}}{\sum_{j=1}^{N}m_{ij,N}}|x_{k,N}\beta_0|\right]\right\} \\
&\leqslant C_1C_2 \sup_{i,N}\left\{\sum_{k=1}^{N}\frac{m_{ik,N}}{\sum_{j=1}^{N}m_{ij,N}}\mathrm{E}\exp\left[2\gamma(1+\zeta)(\sum_{j=1}^{N}m_{ij,N})|x_{k,N}\beta_0|\right]\right\}^{1/2} \\
&\leqslant C_1C_2 \sup_{i,N}\mathrm{E}^{1/2}\left\{\exp\left[\frac{2\gamma(1+\zeta)|x_{i,N}\beta_0|}{1-\sup_N||\lambda_0 W_N||_\infty}\right]\right\} < \infty,
\end{aligned}
$$

where the second inequality is built on the Cauchy–Schwarz inequality, the third and the sixth inequalities are based on the general inequality of arithmetic and geometric means, the fourth inequality originates from $\sum_{k=1}^{N}|\delta_{ik}-\lambda w_{ik,N}| \leqslant 1+\zeta$ and Eq. (C.5), and the fifth inequality is by Minkowski's inequality. □

Proof of Lemma 5: Let $A = \{i \in \{1,2,\cdots,N\} : y_{i,N} > 0\}$ be the set of indexes under which $y_{i,N} > 0$ and $1(A)$ be the event A's indicator. As Y_n is a random vector, each of its realization gives a pattern of zero and positive observations for all its components. Each such a pattern gives an A. Thus, A represents a regime, and $1(A)$ can be interpreted as a regime indicator. For each A, we separate Y_N into two subvectors $Y_{1,N}$, whose elements are all zeros, and

$Y_{2,N}$, whose elements are all positive. Similarly, $Y_N^{*'} = (Y_{1,N}^{*'}, Y_{2,N}^{*'})$. After a proper permutation, $W_N = \begin{pmatrix} W_{11,AN} & W_{12,AN} \\ W_{21,AN} & W_{22,AN} \end{pmatrix}$, so that $Y_{1,N}^* = \lambda_0 W_{12,AN} Y_{2,N} + X_{1,N}\beta_0 + \epsilon_{1,N}$ and $Y_{2,N}^* = Y_{2,N} = \lambda_0 W_{22,AN} Y_{2,N} + X_{2,N}\beta_0 + \epsilon_{2,N}$. Hence $Y_{2,N}^* = (I_{|A|} - \lambda_0 W_{22,AN})^{-1}(X_{2,N}\beta_0 + \epsilon_{2,N})$, where $|A|$ is the cardinality of A.

Next, we calculate the marginal density function $f(y_{i,N}^*)$. As the range of $y_{i,N}^*$ is $(-\infty, +\infty)$, $y_{i,N}^*$ is either strictly positive or negative. When $y_{i,N}^* > 0$, there are 2^{N-1} possible different A's with $i \in A \subset \{1, 2, \cdots, N\}$. On each A, denote the density form at $y_{i,N}^*$ as $f_A(y_{i,N}^*)$.

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

where "$-i$" means the rest $(N-1)$ elements without i. Because the integral of a conditional density function is 1, $\sum_{A \subset \{1,2,\cdots,N\}:\, i \in A} \int f_A(Y_{-i,N}^* | y_{i,N}^*) dY_{-i,N}^* = 1$. Therefore, if we can show that $f_A(y_{i,N}^*)$ is uniformly bounded, then $f(y_{i,N}^*)$ is uniformly bounded on $y_{i,N}^* > 0$. Denote $b_{ij,AN} = ((I_{|A|} - \lambda_0 W_{22,AN})^{-1})_{ij}$, then $y_{i,N}^* = \sum_{j=1}^{|A|} b_{ij,AN}((X_2)_{j,N}\beta_0 + (\epsilon_2)_{j,N})$. Without loss of generality, assume $b_{ij,AN} \neq 0$. The density of $b_{ij,AN}((X_2)_{j,N}\beta_0 + (\epsilon_2)_{j,N})$ is $\tilde{f}_{ij,A}(y) = f_0(y/b_{ij,AN} - (X_2)_{j,N}\beta_0)/|b_{ij,AN}|$. Then $f_A(y_{i,N}^*) = (\tilde{f}_{i1,A} * \tilde{f}_{i2,A} * \cdots * \tilde{f}_{i|A|,A})(y_{i,N}^*)$, where $f * g$ is the convolution of f and g. From Young's inequality in Folland (1999), $||f * g||_\infty \leqslant ||f||_\infty ||g||_1$, where $||f||_\infty \equiv \operatorname{ess\,sup}_x |f(x)|$ and $||g||_1 \equiv \int |g(x)| dx$. Thus,

$$\begin{aligned} &||\tilde{f}_{i1,A} * \tilde{f}_{i2,A} * \cdots * \tilde{f}_{i|A|,A}||_\infty \leqslant ||\tilde{f}_{i1,A} * \tilde{f}_{i2,A} * \cdots * \tilde{f}_{i|A|-1,A}||_\infty ||\tilde{f}_{i|A|,A}||_1 \\ \leqslant &||\tilde{f}_{i1,A}||_1 ||\tilde{f}_{i2,A}||_1 \cdots ||\tilde{f}_{i(i-1),A}||_1 ||\tilde{f}_{ii,A}||_\infty ||\tilde{f}_{i(i+1),A}||_1 \cdots ||\tilde{f}_{i|A|,A}||_1 = ||\tilde{f}_{ii,A}||_\infty. \end{aligned}$$

Consequently, it remains to show that $\sup_{i,N,A} ||\tilde{f}_{ii,A}||_\infty < \infty$. Since $\int f_0(\epsilon) d\epsilon = 1$ and $f_0 \in C^1(\mathbb{R})$, $\sup_{\epsilon \in \mathbb{R}} f_0(\epsilon) < \infty$. So it is sufficient to show that $\inf_{i,N,A} |b_{ii,AN}| > 0$. This holds by Assumptions 2 and 9 as follows. (1) When $\lambda_0 \geqslant 0$, $b_{ii,AN} = ((I_{|A|} - \lambda_0 W_{22,AN})^{-1})_{ii} = 1 + \sum_{j=1}^\infty (\lambda_0 W_{22,AN})_{ii}^j \geqslant 1$. (2) When $\lambda_0 < 0$, because $(W_{22,AN})_{ii} \equiv 0$, we have $b_{ii,AN} = 1 + \sum_{j=2}^\infty (\lambda_0 W_{22,AN})_{ii}^j \geqslant 1 + \sum_{j=1}^\infty (\lambda_0 W_{22,AN})_{ii}^{2j+1} \geqslant 1 - \sum_{j=1}^\infty ||\lambda_0 W_N||_\infty^{2j+1} = 1 - ||\lambda_0 W_N||^3/(1 - ||\lambda_0 W_N||_\infty^2) > 0$, where the last inequality is implied by $||\lambda_0 W_N||_\infty < 0.7548$. (3) When W_N is symmetric or row-normalized from a symmetric matrix, $W_{22,AN} = D_{|A|} W_{22,AN}^*$, where $W_{22,AN}^* = (W_{22,AN}^*)'$ and $D_{|A|}$ is a positive definite diagonal matrix, because we do not consider the rows with all entries zero. By symmetry,

$D_{|A|}^{1/2}W_{22,AN}^{*}D_{|A|}^{1/2} = P_{|A|}\Lambda_{|A|}P_{|A|}'$, where $P_{|A|}$ is a real orthogonal matrix and $\Lambda_{|A|} = (\lambda_1, \cdots, \lambda_{|A|})$ is a diagonal real eigenvalues matrix. Then, $W_{22,AN} = D_{|A|}W_{22,AN}^{*} = D_{|A|}^{1/2}D_{|A|}^{1/2}W_{22,AN}^{*}D_{|A|}^{1/2}D_{|A|}^{-1/2} = D_{|A|}^{1/2}P_{|A|}\Lambda_{|A|}P_{|A|}'D_{|A|}^{-1/2} = (D_{|A|}^{1/2}P_{|A|})\Lambda_{|A|}(D_{|A|}^{1/2}P_{|A|})^{-1}$, and furthermore,

$$
\begin{aligned}
&(I_{|A|} - \lambda_0 W_{22,AN})^{-1} = [I_{|A|} - (D_{|A|}^{1/2}P_{|A|})\lambda_0\Lambda_{|A|}(D_{|A|}^{1/2}P_{|A|})^{-1}]^{-1} \\
=&D_{|A|}^{1/2}P_{|A|}(I_{|A|} - \lambda_0\Lambda_{|A|})^{-1}P_{|A|}'D_{|A|}^{-1/2}.
\end{aligned}
$$

$b_{ii,AN} = ((I_{|A|}-\lambda_0 W_{22,AN})^{-1})_{ii} = [P_{|A|}(I_{|A|}-\lambda_0\Lambda_{|A|})^{-1}P_{|A|}']_{ii} \geqslant (1-\min_k|\lambda_0\lambda_k|)^{-1}\sum_{j=1}^{|A|}(P_{|A|})_{ij}^2 \geqslant (1+\zeta)^{-1}$, because $|\lambda_0\lambda_i| \leqslant ||\lambda_0 W_N||_\infty \leqslant \zeta$, by the spectral radius theorem. (4) When W_N is a lower or upper triangular matrix, $(I_{|A|} - \lambda_0 W_{22,AN})^{-1}$ is also lower or upper triangular, and $b_{ii,AN} = ((I_{|A|} - \lambda_0 W_{22,AN})^{-1})_{ii} = \sum_{l=0}^{\infty}[(\lambda_0 W_{22,AN})^l]_{ii} = 1$.

When $y_{i,N}^{*} < 0$, there are 2^{N-1} possible different A's where $A \subset \{1, 2, \cdots, N\}\backslash\{i\}$. When $A = \emptyset$, $y_{j,N} = 0$ for all j's, $Y_N^{*} = X_N\beta_0 + \epsilon_N$, thus, the relevant density for $y_{i,n}^{*}$ takes the form as the density of f_0. When $A \neq \emptyset$, because $Y_{2,N} = (I_{|A|} - \lambda W_{22,AN})^{-1}(X_{2,N}\beta_0 + \epsilon_{2,N})$,

$$
\begin{aligned}
&Y_{1,N}^{*} = \lambda_0 W_{12,AN}Y_{2,N} + X_{1,N}\beta_0 + \epsilon_{1,N} \\
=&\lambda_0 W_{12,AN}(I_{|A|} - \lambda_0 W_{22,AN})^{-1}(X_{2,N}\beta_0 + \epsilon_{2,N}) + X_{1,N}\beta_0 + \epsilon_{1,N} \\
=&\lambda_0 W_{12,AN}(I_{|A|} - \lambda_0 W_{22,AN})^{-1}X_{2,N}\beta_0 + X_{1,N}\beta_0 + [\lambda_0 W_{12,AN}(I_{|A|} - \lambda_0 W_{22,AN})^{-1}\epsilon_{2,N} + \epsilon_{1,N}].
\end{aligned}
$$

Notice that $\epsilon_{1,N}$ is independent of $\epsilon_{2,N}$. So convolution formula can be extended to include each of the additional components of $\epsilon_{1,N}$, and, once again, Young's inequality implies that $||f_A(y_{i,N}^{*})||_\infty \leqslant ||f_0||_\infty$. □

C.2. The Proof for Section 3

For the proof of consistency of the sieve estimator, we shall rely on the exponential inequalities of NED process and the following lemma (Theorem 2.5, White and Wooldridge, 1991) for uniform convergence.

Lemma C.1. *Let $(\Omega, \mathfrak{F}, P)$ be a complete probability space, let (Θ, ρ) be a metric space, and let $\{\Theta_n\}$ be a sequence of compact subsets of Θ. Let $\{s_{nt} : \Omega \times \Theta_n \to \bar{\mathbb{R}}, n, t = 1, 2, \cdots\}$ and $\{m_{nt} : \Omega \times \Theta_n \to \bar{\mathbb{R}}^{+}, n, t = 1, 2, \cdots\}$ be double arrays of functions such that for each $\theta \in \Theta_n$, $s_{nt}(\theta) \equiv s_{nt}(\cdot, \theta)$ and*

$m_{nt}(\theta) \equiv m_{nt}(\cdot,\theta)$ are measurable-$\mathfrak{F}/\bar{\mathfrak{B}}$. Suppose there exists a sequence $\{d_n : \Theta_n \to \mathbb{R}^+\}$ and a constant $\lambda > 0$ such that for each θ in Θ_n, $|s_{nt}(\theta^0) - s_{nt}(\theta)| < m_{nt}(\theta)\rho(\theta^0,\theta)^\lambda$ for all θ^0 in $\eta_n(\theta) \equiv \{\theta^0 \in \Theta_n : \rho(\theta^0,\theta) < d_n(\theta)\}$. Let $M_n \equiv \sup_{\theta\in\Theta_n}\sum_{t=1}^n \mathrm{E}\, m_{nt}(\theta)$. Suppose further that there exist functions $\gamma_n^s : \Theta_n \times \mathbb{R}^+ \to \mathbb{R}^+$ and $\gamma_n^m : \Theta_n \times \mathbb{R}^+ \to \mathbb{R}^+$ such that for each $\theta \in \Theta_n$, $P\left[|\sum_{t=1}^n[s_{nt}(\theta) - \mathrm{E}\, s_{nt}(\theta)]| > \zeta\right] \leqslant \gamma_n^s(\theta,\zeta)$ and $P\left[|\sum_{t=1}^n[m_{nt}(\theta) - \mathrm{E}\, m_{nt}(\theta)]| > \zeta\right] \leqslant \gamma_n^m(\theta,\zeta)$. Define $\Gamma_n^s(\zeta) \equiv \sup_{\theta\in\Theta_n}\gamma_n^s(\theta,\zeta)$, $\Gamma_n^m(\zeta) \equiv \sup_{\theta\in\Theta_n}\gamma_n^m(\theta,\zeta)$ for all $\zeta \in \mathbb{R}^+$, $n = 1, 2, \cdots$.

Let $H_n(\epsilon)$ be the metric entropy of Θ_n, and let $G_n(\epsilon) \equiv \exp H_n(\epsilon)$; that is, G_n is the smallest number of open sets of radius ϵ that cover θ_n. Let $a_n = O(M_n \inf_{\theta\in\Theta_n} d_n(\theta)^\lambda)$. Then for all $\epsilon > 0$ and all n sufficiently large

$$P\left[\sup_{\theta\in\Theta_n}|S_n(\theta) - \mathrm{E}\, S_n(\theta)]| > \epsilon a_n\right] \leqslant G_n\left((\epsilon a_n/6M_n)^{1/\lambda}\right)\left[\Gamma_n^m(2M_n) + \Gamma_n^s(\epsilon a_n/3)\right],$$

where $S_n(\theta) \equiv \sum_{t=1}^n s_{nt}(\theta)$. If for all $\epsilon > 0$, $G_n\left((\epsilon a_n/6M_n)^{1/\lambda}\right) = o(\min[\Gamma_n^m(2M_n)^{-1}, \Gamma_n^s(\epsilon a_n/3)^{-1}])$, then $P\left[\sup_{\theta\in\Theta_n}|S_n(\theta) - \mathrm{E}\, S_n(\theta)]| > \epsilon a_n\right] \to 0$ as $n \to \infty$.

The following two lemmas are useful for our proof.

Lemma C.2. *For any $0 < I \in \mathbb{Z}$, $|\prod_{i=1}^I f_i(x_1) - \prod_{i=1}^I f_i(x_2)| \leqslant |[f_1(x_1) - f_1(x_2)]\prod_{i=2}^I f_i(x_1)| + \sum_{j=2}^{I-1}|[\prod_{i=1}^{j-1} f_i(x_2)][f_j(x_1) - f_j(x_2)][\prod_{i=j+1}^I f_i(x_1)]| + |[\prod_{i=1}^{I-1} f_i(x_2)][f_I(x_1) - f_I(x_2)]|$.*

Proof of Lemma C.2:

$$\begin{aligned}
&|\prod_{i=1}^I f_i(x_1) - \prod_{i=1}^I f_i(x_2)| = |[f_1(x_1) - f_1(x_2)]\prod_{i=2}^I f_i(x_1) \\
&\quad + \sum_{j=2}^{I-1}[\prod_{i=1}^{j-1} f_i(x_2)][f_j(x_1) - f_j(x_2)][\prod_{i=j+1}^I f_i(x_1)] + [\prod_{i=1}^{I-1} f_i(x_2)][f_I(x_1) - f_I(x_2)]| \\
&\leqslant \left|[f_1(x_1) - f_1(x_2)]\prod_{i=2}^I f_i(x_1)\right| + \sum_{j=2}^{I-1}\left|[\prod_{i=1}^{j-1} f_i(x_2)][f_j(x_1) - f_j(x_2)][\prod_{i=j+1}^I f_i(x_1)]\right| \\
&\quad + \left|[\prod_{i=1}^{I-1} f_i(x_2)][f_I(x_1) - f_I(x_2)]\right|.
\end{aligned}$$

□

Lemma C.3. (*Lemma A.2 in XL, 2014*) *If, for all i and N, $||Y_{i,N}||_{L^{2r}} \leqslant \Delta < \infty$ and $||Z_{i,N}||_{L^{2r}} \leqslant \Delta < \infty$ for some $r > 2$, $||Y_{i,N} - \mathrm{E}[Y_{i,N}|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant d_{i,YN}\psi(s)$ and $||Z_{i,N} - \mathrm{E}[Z_{i,N}|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant$*

$d_{i,ZN}\psi(s)$, then $||Y_{i,N}Z_{i,N}-\mathrm{E}[Y_{i,N}Z_{i,N}|\mathcal{F}_{i,N}(s)]||_{L^2}\leqslant d_{i,N}\tilde{\psi}(s)$, 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 UG L_2-NED, then $\{Y_{i,n}Z_{i,n}\}$ is UG L_2-NED.

Proof of Theorem 1: Under Assumption 12, because $\cup_{n=K^0+2}^{\infty}\Theta_n$ is dense in Θ and $Q_N(\theta)$ is continuous on Θ_n, Proposition 2.4 in White and Wooldridge (1991) implies that $\{\frac{1}{N}Q_N(\theta)\}_{N=1}^{\infty}$ is identifiably unique. With Corollary 2.3 in White and Wooldridge (1991), it suffices to show the uniform convergence in probability: $\sup_{\theta\in\Theta_n}\frac{1}{N}|\ln L_N(\theta)-Q_N(\theta)|=o_p(1)$. There are four terms in the log-likelihood function (Eq. (3)). The proof of the uniform convergence of the second term on the logarithm of the determinant of the Jacobian matrix is the same as that in XL (2015b). Next, we will show the uniform convergence of the other three terms in Eq. (3).

Uniform Convergence of $\frac{1}{N}\sum_{i=1}^{N}1(y_{i,N}>0)\ln g(z_{i,N}(\lambda,\beta))$

Notice that $\frac{1}{N}\sum_{i=1}^{N}\{1(y_{i,N}>0)\ln g(z_{i,N}(\lambda,\beta))-\mathrm{E}[1(y_{i,N}>0)\ln g(z_{i,N}(\lambda,\beta))]\}$ involves only the finite number of parameters in λ and β but does not contain the sieve parameter δ. To show the uniform convergence, by Theorem 1 in Andrews (1992), with the compactness of parameter space of λ and β, it is sufficient to show the pointwise convergence in probability and stochastic equicontinuity of $\{1(y_{i,N}>0)\ln g(z_{i,N}(\lambda,\beta))-\mathrm{E}[1(y_{i,N}>0)\ln g(z_{i,N}(\lambda,\beta))]\}_{i=1}^{N}$. We first show that $\{1(y_{i,N}>0)\ln g(z_{i,N}(\lambda,\beta))\}_{i=1}^{N}$ is a uniformly L_2-NED random field and uniformly L_p bounded for some $p>1$. Because $|d\ln g(x)/dx|<c(|x|+1)$ (Assumption 13) and $\{z_{i,N}(\lambda,\beta)\}_{i=1}^{N}$ is uniformly L_p bounded for any $p>1$ (Lemma 3), by Lemma 4, $\{1(y_{i,N}>0)\ln g(z_{i,N}(\lambda,\beta))\}_{i=1}^{N}$ is also both uniformly L_p bounded and UG L_2-NED in λ, β, i and N. With Assumptions 6 and 14, the pointwise convergence in probability is a result of LLN for spatial NED process of Theorem 1 in JP (2012). By Lemma 1 in Andrews (1992), the stochastic equicontinuity of $\{1(y_{i,N}>0)\ln g(z_{i,N}(\lambda,\beta))-\mathrm{E}[1(y_{i,N}>0)\ln g(z_{i,N}(\lambda_1,\beta_1))]\}_{i=1}^{N}$ is implied by the uniform L_p

boundedness of $|w_{i\cdot,N}Y_N|$ and $x_{i,N}$ in the linear function $z_{i,N}(\lambda,\beta)$, and

$$\begin{aligned}
&|1(y_{i,N}>0)\ln g(z_{i,N}(\lambda_1,\beta_1))-1(y_{i,N}>0)\ln g(z_{i,N}(\lambda_2,\beta_2))|\\
\leqslant&|\ln g(z_{i,N}(\lambda_1,\beta_1))-\ln g(z_{i,N}(\lambda_2,\beta_2))|\\
\leqslant&c[|z_{i,N}(\lambda_1,\beta_1)|+|z_{i,N}(\lambda_2,\beta_2)|+1]\cdot|z_{i,N}(\lambda_1,\beta_1)-z_{i,N}(\lambda_2,\beta_2)|\\
\leqslant&c[|z_{i,N}(\lambda_1,\beta_1)|+|z_{i,N}(\lambda_2,\beta_2)|+1][|w_{i\cdot,N}Y_N|+\sum_{k=1}^{K^0}|x_{ik,N}|]\cdot[|\lambda_1-\lambda_2|+\sum_{k=1}^{K^0}|\beta_{1k}-\beta_{2k}|].
\end{aligned}$$

Uniform Convergence of $\frac{1}{N}\sum_{i=1}^{N}1(y_{i,N}>0)\ln h(G(z_{i,N}(\lambda,\beta))|\delta)$ for $\theta\in\Theta_n$

Denote $s_{i,N}(\theta)=1(y_{i,N}>0)\ln h(G(z_{i,N}(\lambda,\beta))|\delta)$ and we use Lemma C.1 to establish the uniform convergence. We will show the conclusion in six steps. Denote $\theta_1=(\lambda_1,\beta_1',\delta_1)$ and $\theta_2=(\lambda_2,\beta_2',\delta_2)$.

First, calculate the Lipschitz coefficient of this term.

$$\begin{aligned}
|s_{i,N}(\theta_1)-s_{i,N}(\theta_2)|\leqslant&|\ln h(G(z_{i,N}(\lambda_1,\beta_1))|\delta_1)-\ln h(G(z_{i,N}(\lambda_1,\beta_1))|\delta_2)|\\
&+|\ln h(G(z_{i,N}(\lambda_1,\beta_1))|\delta_2)-\ln h(G(z_{i,N}(\lambda_2,\beta_2))|\delta_2)|.
\end{aligned}\tag{C.6}$$

For any $0<a\leqslant b\leqslant c$, $|\ln b-\ln c|\leqslant(c-b)/a$. For the first term on the right hand side of Eq. (C.6), since $h(u|\delta)\geqslant\epsilon_0$ and $\theta\in\Theta_n$,

$$\begin{aligned}
&|\ln h(G(z_{i,N}(\lambda_1,\beta_1))|\delta_1)-\ln h(G(z_{i,N}(\lambda_1,\beta_1))|\delta_2)|\\
\leqslant&\epsilon_0^{-1}|h(G(z_{i,N}(\lambda_1,\beta_1))|\delta_1)-h(G(z_{i,N}(\lambda_1,\beta_1))|\delta_2)|\\
\leqslant&\epsilon_0^{-1}[2M_N(1+\sqrt{2}M_N)^2+2(\sqrt{2}+2M_N)]\cdot||\delta_1-\delta_2||_0,
\end{aligned}$$

where the second inequality is by Lemma B.1 (3). For the second term on the right hand side of

Eq. (C.6), by Assumption 13(2) and Lemma B.1(7),

$$\begin{aligned}
&|\ln h(G(z_{i,N}(\lambda_1,\beta_1))|\delta_2) - \ln h(G(z_{i,N}(\lambda_2,\beta_2))|\delta_2)| \\
\leqslant&\pi M_N(2/\epsilon_0)^{1/2}|G(z_{i,N}(\lambda_1,\beta_1)) - G(z_{i,N}(\lambda_2,\beta_2))| \\
\leqslant&\pi M_N(2/\epsilon_0)^{1/2}C_g|z_{i,N}(\lambda_1,\beta_1) - z_{i,N}(\lambda_2,\beta_2)| \\
\leqslant&\pi M_N(2/\epsilon_0)^{1/2}C_g[|w_{i\cdot,N}Y_N|\cdot|\lambda_1-\lambda_2| + \sum_{k=1}^{K^0}|x_{ik,N}|\cdot|\beta_{1k}-\beta_{2k}|].
\end{aligned} \tag{C.7}$$

Then, $|s_{i,N}(\theta_1) - s_{i,N}(\theta_2)| \leqslant m_{i,N}\cdot[|\lambda_1-\lambda_2| + \sum_{k=1}^{K^0}|\beta_{1k}-\beta_{2k}| + ||\delta_1-\delta_2||_0] = m_{i,N}||\theta_1-\theta_2||_0$, where $m_{i,N} \equiv C_1 M_N(|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0}|x_{ik,N}| + M_N^2)$ for some constant $C_1>0$.

Second, calculate an upper bound of $\sum_{i=1}^N \mathrm{E}\, m_{i,N}$, denoted as $\overline{M_N}$. Since $\sup_{i,N}\mathrm{E}(|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0}|x_{ik,N}|) < \infty$, we can take $\overline{M_N} \equiv NC_2M_N^3$ for some $C_2>0$.

Third, show that $\{s_{i,N}(\theta)\}_{i=1}^N$ is a UG L_2-NED such that the exponential inequalities for NED random fields in Appendix A can be utilized. By Eq. (C.7),

$$\begin{aligned}
&\sup_{i,\theta\in\Theta_n}||\ln h(G(z_{i,N}(\lambda,\beta))|\delta) - \mathrm{E}[\ln h(G(z_{i,N}(\lambda,\beta))|\delta)|\mathcal{F}_{i,N}(s)]||_{L^2} \\
\leqslant&\sup_{i,\theta\in\Theta_n}||\ln h\,[G(z_{i,N}(\lambda,\beta))|\delta] - \ln h\,\{G\,(\mathrm{E}[z_{i,N}(\lambda,\beta)|\mathcal{F}_{i,N}(s)]|\,\delta)\}\,||_{L^2} \\
\leqslant&\sup_{i,\theta\in\Theta_n}\pi M_N(2/\epsilon_0)^{1/2}C_g||z_{i,N}(\lambda,\beta) - \mathrm{E}[z_{i,N}(\lambda,\beta)|\mathcal{F}_{i,N}(s)]||_{L^2} \\
=&\pi M_N(2/\epsilon_0)^{1/2}C_gC_z(\zeta^{1/\bar{d}_0})^s.
\end{aligned} \tag{C.8}$$

From Lemma B.1 (1), $\ln\epsilon_0 \leqslant \ln h(u|\delta) \leqslant 2\ln(1+\sqrt{2}M_N)$. Thus, $|\ln h(u|\delta)| \leqslant \max(2\ln(1+\sqrt{2}M_N), \ln\epsilon_0^{-1})$. By Corollary 1, Lemma C.2, and Eq. (C.8), for some constant $C_3>0$, we have

$$\begin{aligned}
&\sup_{i,\theta\in\Theta}||s_{i,N}(\theta) - \mathrm{E}[s_{i,N}(\theta)|\mathcal{F}_{i,N}(s)]||_{L^2} \\
\leqslant&[\max(2\ln(1+\sqrt{2}M_N), \ln\epsilon_0^{-1})C_{1(y>0)}\zeta^{s/3\bar{d}_0} + \pi M_N\sqrt{\frac{2}{\epsilon_0}}C_gC_z(\zeta^{1/\bar{d}_0})^s] \leqslant C_3M_N\zeta^{s/3\bar{d}_0}.
\end{aligned}$$

Fourth, we calculate the exponential rates for $m_{i,N}$ and $s_{i,N}$ by exponential inequalities in Appendix A. With Assumptions 6 and 7, Lemma 3 holds, and the exponential inequality in Theorem A.2 with $\alpha=1$ is applicable to the UG L_2-NED $\{|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0}|x_{ik,N}|\}_{i=1}^N$: for any $A>0$, there exist some finite positive constants C_{41} and C_{42} such that

$$\begin{aligned}
&P\left(\left|\sum_{i=1}^{N}(m_{i,N}-\mathrm{E}\,m_{i,N})\right|\geqslant A\right)\\
=&P\left(\left|\sum_{i=1}^{N}[|w_{i\cdot,N}Y_N|+\sum_{k=1}^{K^0}|x_{ik,N}|-\mathrm{E}\,|w_{i\cdot,N}Y_N|-\mathrm{E}\sum_{k=1}^{K^0}|x_{ik,N}|]\right|\geqslant\frac{A}{C_1M_N}\right)\\
\leqslant&C_{41}(\frac{NM_N}{A}+1)\exp\left[-C_{42}(\frac{A^2}{NM_N^2})^{1/(2d+4)}\right]\equiv\Gamma_N^m(A).
\end{aligned}$$

Because $\sup_{i,N,\theta\in\Theta_n}|s_{i,N}(\theta)|\leqslant\max(\ln(1+\sqrt{2}M_N),-\ln\epsilon_0)$, the exponential inequality for bounded NED random field in Corollary A.3 is applicable: for any $A>0$, there exist finite positive constants C_{51} and C_{52} such that, for any $\theta\in\Theta_n$,

$$P\left(\left|\sum_{i=1}^{N}[s_{i,N}(\theta)-\mathrm{E}\,s_{i,N}(\theta)]\right|\geqslant A\right)\leqslant C_{51}\exp\left[-C_{52}\left(\frac{A^2}{N\ln^2M_N}\right)^{1/(2d+2)}\right]\equiv\Gamma_N^s(A).$$

Fifth, for each sample size N, find a covering number of $\Theta_n\equiv\{(\lambda,\beta',\delta)\in\Theta:|\lambda|\leqslant\lambda_m,\,|\beta_k|\leqslant\beta_{km},\,1\leqslant k\leqslant K^0,||\delta||_{l_0}\leqslant M_N,\delta_i=0,\forall i\geqslant n-K^0\}$. Let $\tilde{\delta_j}\equiv j^{l_0}\delta_j$. Then $\widetilde{\Theta_n}=\{(\lambda,\beta',\tilde{\delta}):|\lambda|\leqslant\lambda_m,\,|\beta_k|\leqslant\beta_{km},\,1\leqslant k\leqslant K^0,\,\sum_{j=1}^{n-K^0-1}|\tilde{\delta_j}|\leqslant M_N,\tilde{\delta}_i=0,\forall i>n-K^0\}$. Clearly, $(\Theta_n,||(\lambda,\beta,\delta)||_{l_0})$ and $(\widetilde{\Theta_n},\,||(\lambda,\beta,\tilde{\delta})||_0)$ have the same covering numbers, which is less than or equal to that of

$$\Theta_n^*=\{(\lambda,\beta_1,\cdots,\beta_{K^0},\tilde{\delta_1},\ldots,\tilde{\delta}_{n-K^0-1}):\,|\lambda|+\sum_{k=1}^{K^0}|\beta_k|+\sum_{j=1}^{n-K^0-1}|\tilde{\delta_j}|\leqslant\lambda_m+\sum_{k=1}^{K^0}\beta_{km}+M_N\},$$

which is $G_n^*(r)\leqslant[2r^{-1}(\lambda_m+\sum_{k=1}^{K^0}\beta_{km}+M_N)+1]^n$, where r is the radius of a ball with $r\leqslant\lambda_m+\sum_{k=1}^{K^0}\beta_{km}+M_N$, by example 12.3 in lecture 12 in Panchenko (2007). Thus, the covering number of Θ_n, $G_n(r)\leqslant[2r^{-1}(\lambda_m+\sum_{k=1}^{K^0}\beta_{km}+M_N)+1]^n$.

Sixth, by Theorem 2.5 in White and Wooldridge (1991), for all $\epsilon>0$ and all N large enough,

$$\begin{aligned}
&P(\sup_{\theta\in\Theta_n}\sum_{i=1}^{N}[s_{i,N}(\theta)-\mathrm{E}\,s_{i,N}(\theta)]\geqslant N\epsilon)\leqslant G_n(\frac{N\epsilon}{6\overline{M_N}})[\Gamma_N^m(2\overline{M_N})+\Gamma_N^s(\epsilon N/3)]\\
\leqslant&[\frac{12C_2M_N^3(\lambda_m+\sum_{k=1}^{K^0}\beta_{km}+M_N)}{\epsilon}+1]^n\{C_{41}(\frac{1}{2C_2M_N^2}+1)\cdot\\
&\exp\left[-C_{42}(4C_2^2NM_N^4)^{1/(2d+4)}\right]+C_{51}\exp[-C_{52}(\frac{\epsilon^2N}{9\ln^2M_N})^{1/(2d+2)}]\}.
\end{aligned}$$

The right hand side of the above inequality is $o(1)$ when $\lim_{N\to\infty}[n\ln(M_N^4)]/(NM_N^4)^{1/(2d+4)} = 0$ and $\lim_{N\to\infty}[n\ln(M_N^4)]/(N/\ln^2 M_N)^{1/(2d+2)} = 0$, i.e, $\lim_{N\to\infty}(n^{2d+4}\ln^{2d+4} M_N)/(NM_N^4) = 0$ and $\lim_{N\to\infty}(n^{2d+2}\ln^{2d+4} M_N)/N = 0$, which are satisfied under Eq. (8). Then, the uniform convergence is established.

Uniform Convergence of $\frac{1}{N}\sum_{i=1}^N 1(y_{i,N}=0)\ln H(G(z_{i,N}(\lambda,\beta))|\delta)$

Similarly, now let $s_{i,N}(\theta) = 1(y_{i,N}>0)\ln H(G(z_{i,N}(\lambda,\beta))|\delta)$. First, study the Lipschitz coefficient of $s_{i,N}(\theta)$. $|s_{i,N}(\lambda_1,\beta_1,\delta)-s_{i,N}(\lambda_2,\beta_2,\delta)| \leqslant \epsilon_0^{-1}(1+\sqrt{2}M_N)^2 c[|z_{i,N}(\lambda_1,\beta_1)|+|z_{i,N}(\lambda_2,\beta_2)|+1]\cdot|(\lambda_1-\lambda_2)w_{i\cdot,N}Y_N + x_{i,N}(\beta_1-\beta_2)|$, because

$$\sup_{||\delta||_1\leqslant M_N}|\frac{\partial \ln H(G(x)|\delta)}{\partial x}| = \sup_{||\delta||_1\leqslant M_N}\frac{h(G(x)|\delta)}{H(G(x)|\delta)/G(x)}\frac{g(x)}{G(x)} \leqslant \epsilon_0^{-1}(1+\sqrt{2}M_N)^2 c(|x|+1), \quad \text{(C.9)}$$

In addition, for $||\delta_1||_1 \leqslant M_N$ and $||\delta_2||_1 \leqslant M_N$,

$$\begin{aligned}
&|\ln H(u|\delta_1) - \ln H(u|\delta_2)| \leqslant \frac{|H(u|\delta_1)-H(u|\delta_2)|}{\epsilon_0 u} \leqslant (\epsilon_0 u)^{-1}\int_0^u |h(v|\delta_1)-h(v|\delta_2)|dv \\
\leqslant&\epsilon_0^{-1}\sup_{v\in[0,1]}|h(v|\delta_1)-h(v|\delta_2)| \leqslant \epsilon_0^{-1}[2M_N(1+\sqrt{2}M_N)^2 + 2(\sqrt{2}+2M_N)]\cdot||\delta_1-\delta_2||_0,
\end{aligned}$$

where the first inequality is built on $H(u|\delta) \geqslant u\epsilon_0$ and the last inequality originates from Lemma B.1 (3). Because $z_{i,N}(\lambda,\beta) = -\lambda w_{i\cdot,N}Y_N - x_{i,N}\beta$ when $y_{i,N}=0$,

$$\begin{aligned}
&|s_{i,N}(\theta_1) - s_{i,N}(\theta_2)| \leqslant |\ln H(G(z_{i,N}(\lambda_1,\beta_1))|\delta_1) - \ln H(G(z_{i,N}(\lambda_2,\beta_2))|\delta_2)| \\
\leqslant&|\ln H(G(z_{i,N}(\lambda_1,\beta_1))|\delta_1) - \ln H(G(z_{i,N}(\lambda_2,\beta_2))|\delta_1)| + \\
&|\ln H(G(z_{i,N}(\lambda_2,\beta_2))|\delta_1) - \ln H(G(z_{i,N}(\lambda_2,\beta_2))|\delta_2)| \\
\leqslant&\epsilon_0^{-1}(1+\sqrt{2}M_N)^2 c[|z_{i,N}(\lambda_1,\beta_1)|+|z_{i,N}(\lambda_2,\beta_2)|+1]\cdot|(\lambda_1-\lambda_2)w_{i\cdot,N}Y_N + x_{i,N}(\beta_1-\beta_2)| \\
&+\epsilon_0^{-1}[2M_N(1+\sqrt{2}M_N)^2 + 2(\sqrt{2}+2M_N)]\cdot||\delta_1-\delta_2||_0 \\
\leqslant&C_6 M_N^2[(|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0}|x_{ik,N}|)^2 + M_N]\cdot||\theta_1-\theta_2||_0
\end{aligned}$$

for some constant $C_6>0$. Thus, we define $m_{i,N} = C_6 M_N^2[(|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0}|x_{ik,N}|)^2 + M_N]$. So $\mathrm{E}\sum_{i=1}^N m_{i,N} \leqslant NC_7 M_N^3 \equiv \widetilde{M_N}$ for some $C_7>0$ and for all N.

Second, establish the NED properties of $\{(|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0}|x_{ik,N}|)^2\}_{i=1}^N$ and $\{s_{i,N}(\theta)\}_{i=1}^N$. By

Eq. (C.9), $|\ln H(G(z_{i,N}(\lambda,\beta))|\delta)| \leqslant |\ln H(G(0)|\delta)| + \epsilon_0^{-1}(1+\sqrt{2}M_N)^2 c(|z_{i,N}(\lambda,\beta)|+1)|z_{i,N}(\lambda,\beta)|$. Because, at $y_{i,N}=0$, $z_{i,N}(\lambda,\beta) = -(\lambda_{i,N}Y_N + x_{i,N}\beta)$, by Lemma 3,

$$\sup_{i,N,\theta\in\Theta_n} \mathrm{E}\exp[\gamma|M_N^{-2}\ln H(G(z_{i,N}(\lambda,\beta))|\delta)|^{1/2}] < \infty \tag{C.10}$$

for some constant $\gamma > 0$. By similar argument for Lemma 3, we have

$$\sup_{i,N} \mathrm{E}\exp[\bar{\gamma}(|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0}|x_{ik,N}|)] < \infty \tag{C.11}$$

for some constant $\bar{\gamma} > 0$. As a result, because $\{z_{i,N}(\lambda,\beta)\}_{i=1}^N$ and $\{w_{i\cdot,N}Y_N\}_{i=1}^N$ are UG L_2-NED, Lemmas 3 and 4 implies that for any $\gamma_1 \in (\frac{1}{3},\frac{1}{2})$, $\{(|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0}|x_{ik,N}|)^2\}_{i=1}^N$ and $\{M_N^{-2}\ln H(G(z_{i,N}(\lambda,\beta))|\delta)\}_{i=1}^N$ are UG L_2-NED, with NED coefficient $\zeta^{\gamma_1 s/\bar{d}_0}$.[5] Hence Lemma C.3 implies that for some $\gamma_2 \in (0,\frac{1}{6})$ and $C_8 > 0$, $||M_N^{-2}s_{i,N}(\theta) - \mathrm{E}[M_N^{-2}s_{i,N}(\theta)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_8\zeta^{s\gamma_2/\bar{d}_0}$.

Third, calculate the exponential inequalities of $m_{i,N}$ and $s_{i,N}(\theta)$. With Eqs. (C.10) and (C.11), Assumption A.3 holds for $(|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0}|x_{ik,N}|)^2$ and $M_N^{-2}s_{i,N}(\theta)$. Accordingly, Theorem A.2 with $\alpha = \frac{1}{2}$ is applicable. For any $A > 0$, there exist positive finite constants C_{91}, C_{92}, $C_{10,1}$ and $C_{10,2}$ such that

$$\begin{aligned}
&P(|\sum_{i=1}^N (m_{i,N} - \mathrm{E}\,m_{i,N})| \geqslant A) \\
=&P\left(\left|\sum_{i=1}^N\left[(|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0}|x_{ik,N}|)^2 - \mathrm{E}(|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0}|x_{ik,N}|)^2\right]\right| \geqslant \frac{A}{C_6M_N^2}\right) \\
\leqslant&C_{91}(\frac{NM_N^2}{A}+1)\exp[-C_{92}(\frac{A^2}{NM_N^4})^{1/(2d+6)}] \equiv \Gamma_N^m(A),
\end{aligned}$$

[5] The scale number in front of the NED coefficient depends on any finite order of moments but not infinite moment, so Lemma 4 can not be used even arbitrarily high moment exits and, therefore, γ_1 can not be taken with the value 0.5.

and

$$P\left(|\sum_{i=1}^{N}(s_{i,N}-\mathrm{E}\,s_{i,N})|\geqslant A\right)=P\left(|\sum_{i=1}^{N}[M_N^{-2}s_{i,N}-\mathrm{E}(M_N^{-2}s_{i,N})]|\geqslant M_N^{-2}A\right)$$
$$\leqslant C_{10,1}(\frac{NM_N^2}{A}+1)\exp[-C_{10,2}(\frac{A^2}{NM_N^4})^{1/(2d+6)}]\equiv\Gamma_N^s(A).$$

Fourth, by Theorem 2.5 in White and Wooldridge (1991) and Eq. (8), for all $\epsilon>0$ and all N large enough,

$$P\left(\sup_{\theta\in\Theta_n}\left|\sum_{i=1}^{N}(s_{i,N}-\mathrm{E}\,s_{i,N})\right|\geqslant N\epsilon\right)\leqslant G_n(\frac{N\epsilon}{6\widetilde{M_N}})[\Gamma_N^m(2\widetilde{M_N})+\Gamma_N^s(\epsilon N/3)]$$
$$\leqslant[12C_7\epsilon^{-1}M_N^3(\lambda_m+\sum_{k=1}^{K^0}\beta_{km}+M_N)+1]^n\cdot\{C_{91}(\frac{1}{2C_7M_N}+1)\exp[-C_{92}(4C_7^2NM_N^2)^{1/(2d+6)}]$$
$$+C_{10,1}(\frac{3M_N^2}{\epsilon}+1)\exp[-C_{10,2}(\frac{N\epsilon^2}{9M_N^4})^{1/(2d+6)}]\}.$$

The right hand side of the above inequality is $o(1)$ when $\lim_{N\to\infty}[n\ln(M_N^4)]/(NM_N^2)^{1/(2d+6)}=0$ and $\lim_{N\to\infty}[n\ln(M_N^4)]/(N/M_N^4)^{1/(2d+6)}=0$, i.e., $\lim_{N\to\infty}(n^{2d+6}\ln^{2d+6}M_N)/NM_N^2=0$ and $\lim_{N\to\infty}(n^{2d+6}M_N^4\ln^{2d+6}M_N)/N=0$, which hold by Eq. (8). Hence, the consistency of the sieve estimator follows. □

D. Derivatives of the log-likelihood function and their properties

The derivatives of the log-likelihood function are

$$\nabla_\lambda L_{i,N}(\theta)=-[1(y_{i,N}=0)\frac{h(G(z_{i,N}(\lambda,\beta))|\delta)}{H(G(z_{i,N}(\lambda,\beta))|\delta)}+1(y_{i,N}>0)\frac{h'(G(z_{i,N}(\lambda,\beta))|\delta)}{h(G(z_{i,N}(\lambda,\beta))|\delta)}]\cdot$$
$$g(z_{i,N}(\lambda,\beta))w_{i\cdot,N}Y_N-1(y_{i,N}>0)\frac{g'(z_{i,N}(\lambda,\beta))}{g(z_{i,N}(\lambda,\beta))}w_{i\cdot,N}Y_N-\sum_{k=1}^{\infty}\lambda^{k-1}(\widetilde{W_N}^k)_{ii}, \tag{D.1}$$

$$\nabla_{\beta_k} L_{i,N}(\theta) = -[1(y_{i,N}=0)\frac{h(G(z_{i,N}(\lambda,\beta))|\delta)}{H(G(z_{i,N}(\lambda,\beta))|\delta)} + 1(y_{i,N}>0)\frac{h'(G(z_{i,N}(\lambda,\beta))|\delta)}{h(G(z_{i,N}(\lambda,\beta))|\delta)}]\cdot$$
$$g(z_{i,N}(\lambda,\beta))x_{ik,N} - 1(y_{i,N}>0)\frac{g'(z_{i,N}(\lambda,\beta))}{g(z_{i,N}(\lambda,\beta))}x_{ik,N}, \tag{D.2}$$

$$\nabla_{\delta_k} L_{i,N}(\theta) = 1(y_{i,N}=0)\frac{\nabla_{\delta_k} H(G(z_{i,N}(\lambda,\beta))|\delta)}{H(G(z_{i,N}(\lambda,\beta))|\delta)} + 1(y_{i,N}>0)\frac{\nabla_{\delta_k} h(G(z_{i,N}(\lambda,\beta))|\delta)}{h(G(z_{i,N}(\lambda,\beta))|\delta)}. \tag{D.3}$$

Lemma D.1. *Under Assumptions 10 and 13, there exists a constant $C>0$ that depends on neither i nor N, such that*

(1) $|\nabla_\lambda L_{i,N}(\theta)| \leqslant C(1+||\delta||_1^2)[1+|z_{i,N}(\lambda,\beta)|]\cdot|w_{i\cdot,N}Y_N| + C$,

(2) $|\nabla_{\beta_k} L_{i,N}(\theta)| \leqslant C(1+||\delta||_1^2)[1+|z_{i,N}(\lambda,\beta)|]\cdot|x_{ik,N}|$

(3) $|\nabla_{\delta_k} L_{i,N}(\theta)| \leqslant C(1+||\delta||_0^2)$.

Proof of Lemma D.1: By Lemmas B.1 (1) and (7), B.2 (1) and Assumption 13, $\frac{h(G(x)|\delta)}{H(G(x)|\delta)}g(x) = \frac{h(G(x)|\delta)}{H(G(x)|\delta)/G(x)}\frac{g(x)}{G(x)} \leqslant \epsilon_0^{-1}(1+\sqrt{2}||\delta||_0)^2 c(|x|+1)$ and $|\frac{h'(G(x)|\delta)}{h(G(x)|\delta)}g(x)| \leqslant C_g\pi||\delta||_1(2/\epsilon_0)^{1/2}$. In addition, $|\sum_{k=1}^\infty \lambda^{k-1}(\widetilde{W_N}^k)_{ii}| \leqslant \lambda_m^{-1}\sum_{k=0}^\infty ||\lambda_m \widetilde{W_N}||_\infty^k \leqslant \lambda_m^{-1}/(1-\zeta)$. Thus, the first two conclusions hold. The bound for $|\nabla_{\delta_k} L_{i,N}(\theta)|$ is by Lemmas B.1 (2) and B.2 (5). □

Denote $\psi_1(u|\delta) = \frac{u}{H(u|\delta)}$. The second order derivatives of the log-likelihood function are:

$$\begin{aligned}
&\nabla_{\lambda,\lambda} \ln L_N(\theta)\\
=&\sum_{i=1}^N \Bigg\{1(y_{i,N}=0)\left[h(u|\delta)\psi_1(u|\delta)\frac{g'(z)}{G(z)} + h'(u|\delta)\psi_1(u|\delta)\frac{g^2(z)}{G(z)} - \left(h(u|\delta)\psi_1(u|\delta)\frac{g(z)}{G(z)}\right)^2\right]\\
&+1(y_{i,N}>0)\left[\frac{h'(u|\delta)}{h(u|\delta)}g'(z) + \frac{h''(u|\delta)}{h(u|\delta)}g^2(z) - \left(\frac{h'(u|\delta)}{h(u|\delta)}g(z)\right)^2 + \left(\frac{g''(z)}{g(z)} - \frac{g'(z)^2}{g(z)^2}\right)\right]\Bigg\}\\
&\cdot(w_{i\cdot,N}Y_N)^2\,\big|_{u=G(z_{i,N}(\lambda,\beta)),z=z_{i,N}(\lambda,\beta)} - \sum_{i=1}^N\sum_{k=1}^\infty k\lambda^{k-1}(\widetilde{W_N}^{k+1})_{ii},
\end{aligned} \tag{D.4}$$

$$\begin{aligned}
&\nabla_{\lambda,\beta_k} \ln L_N(\theta)\\
=&\sum_{i=1}^N \Bigg\{1(y_{i,N}=0)\left[h(u|\delta)\psi_1(u|\delta)\frac{g'(z)}{G(z)} + h'(u|\delta)\psi_1(u|\delta)\frac{g^2(z)}{G(z)} - \left(h(u|\delta)\psi_1(u|\delta)\frac{g(z)}{G(z)}\right)^2\right]\\
&+1(y_{i,N}>0)\left[\frac{h'(u|\delta)}{h(u|\delta)}g'(z) + \frac{h''(u|\delta)}{h(u|\delta)}g^2(z) - \left(\frac{h'(u|\delta)}{h(u|\delta)}g(z)\right)^2 + \left(\frac{g''(z)}{g(z)} - \frac{g'(z)^2}{g(z)^2}\right)\right]\Bigg\}\\
&\cdot x_{ik,N}\cdot w_{i\cdot,N}Y_N\,\big|_{u=G(z_{i,N}(\lambda,\beta)),z=z_{i,N}(\lambda,\beta)}\ ,
\end{aligned} \tag{D.5}$$

$$
\begin{aligned}
&\nabla_{\beta_j,\beta_k} \ln L_N(\theta) \\
=&\sum_{i=1}^{N}\left\{1(y_{i,N}=0)\left[h(u|\delta)\psi_1(u|\delta)\frac{g'(z)}{G(z)}+h'(u|\delta)\psi_1(u|\delta)\frac{g^2(z)}{G(z)}-\left(h(u|\delta)\psi_1(u|\delta)\frac{g(z)}{G(z)}\right)^2\right]\right. \\
&\left.+1(y_{i,N}>0)\left[\frac{h'(u|\delta)}{h(u|\delta)}g'(z)+\frac{h''(u|\delta)}{h(u|\delta)}g^2(z)-\left(\frac{h'(u|\delta)}{h(u|\delta)}g(z)\right)^2+\left(\frac{g''(z)}{g(z)}-\frac{g'(z)^2}{g(z)^2}\right)\right]\right\} \\
&\cdot x_{ik,N}x_{ij,N}\left.\right|_{u=G(z_{i,N}(\lambda,\beta)),z=z_{i,N}(\lambda,\beta)} ,
\end{aligned}
\tag{D.6}
$$

$$
\begin{aligned}
&\nabla_{\lambda,\delta_k} \ln L_N(\theta) \\
=&-\sum_{i=1}^{N}\left\{1(y_{i,N}=0)\left[\nabla_{\delta_k}h(u|\delta)\psi_1(u|\delta)-\frac{\nabla_{\delta_k}H(u|\delta)}{u}h(u|\delta)\psi_1^2(u|\delta)\right]\frac{g(z_{i,N}(\lambda,\beta))}{G(z_{i,N}(\lambda,\beta))}+\right. \\
&\left.1(y_{i,N}>0)\left[\frac{\nabla_{\delta_k}h'(u|\delta)}{h(u|\delta)}-\frac{\nabla_{\delta_k}h(u|\delta)}{h^2(u|\delta)}h'(u|\delta)\right]g(z_{i,N}(\lambda,\beta))\right\}w_{i\cdot,N}Y_N\left.\right|_{u=G(z_{i,N}(\lambda,\beta))} ,
\end{aligned}
\tag{D.7}
$$

$$
\begin{aligned}
&\nabla_{\beta_j,\delta_k} \ln L_N(\theta) \\
=&-\sum_{i=1}^{N}\left\{1(y_{i,N}=0)\left[\nabla_{\delta_k}h(u|\delta)\psi_1(u|\delta)-\frac{\nabla_{\delta_k}H(u|\delta)}{u}h(u|\delta)\psi_1^2(u|\delta)\right]\frac{g(z_{i,N}(\lambda,\beta))}{G(z_{i,N}(\lambda,\beta))}+\right. \\
&\left.1(y_{i,N}>0)\left[\frac{\nabla_{\delta_k}h'(u|\delta)}{h(u|\delta)}-\frac{\nabla_{\delta_k}h(u|\delta)}{h^2(u|\delta)}h'(u|\delta)\right]g(z_{i,N}(\lambda,\beta))\right\}x_{ij,N}\left.\right|_{u=G(z_{i,N}(\lambda,\beta))} ,
\end{aligned}
\tag{D.8}
$$

$$
\begin{aligned}
\nabla_{\delta_j,\delta_k} \ln L_N(\theta)=&\sum_{i=1}^{N}\left\{1(y_{i,N}=0)\left[\frac{\nabla_{\delta_j,\delta_k}H(u|\delta)}{H(u|\delta)}-\frac{\nabla_{\delta_j}H(u|\delta)}{H(u|\delta)}\frac{\nabla_{\delta_k}H(u|\delta)}{H(u|\delta)}\right]+\right. \\
&\left.1(y_{i,N}>0)\left[\frac{\nabla_{\delta_j,\delta_k}h(u|\delta)}{h(u|\delta)}-\frac{\nabla_{\delta_j}h(u|\delta)\nabla_{\delta_k}h(u|\delta)}{h^2(u|\delta)}\right]\right\}\left.\right|_{u=G(z_{i,N}(\lambda,\beta))} .
\end{aligned}
\tag{D.9}
$$

Lemma D.2. *Under Assumptions 10 and 13, there exists a constant C that does not depend on i, N, j, k or θ, such that*

(1) $|\nabla_{\lambda,\lambda}L_{i,N}(\theta)| \leqslant C(1+||\delta||_2^4)[1+z_{i,N}^2(\lambda,\beta)](w_{i\cdot,N}Y_N)^2+C$;

$|\nabla_{\lambda,\beta_k}L_{i,N}(\theta)| \leqslant C(1+||\delta||_2^4)[1+z_{i,N}^2(\lambda,\beta)]|x_{ik,N}\cdot w_{i\cdot,N}Y_N|$;

$|\nabla_{\beta_j,\beta_k}L_{i,N}(\theta)| \leqslant C(1+||\delta||_2^4)[1+z_{i,N}^2(\lambda,\beta)]|x_{ik,N}x_{ij,N}|$;

$|\nabla_{\lambda,\delta_k}L_{i,N}(\theta)| \leqslant Ck(1+||\delta||_1^4)(1+|z_{i,N}(\lambda,\beta)|)|w_{i\cdot,N}Y_N|$;

$|\nabla_{\beta_j,\delta_k}L_{i,N}(\theta)| \leqslant Ck(1+||\delta||_1^4)(1+|z_{i,N}(\lambda,\beta)|)|x_{ij,N}|$;

$|\nabla_{\delta_j,\delta_k} L_{i,N}(\theta)| \leqslant C(1+||\delta||_0^4)$.

(2) *Let* $\theta^1 = (\lambda_1, \beta^1, \delta^1)$ *and* $\theta^2 = (\lambda_2, \beta^2, \delta^2)$ *be such that* $||\theta^1 - \theta^2||_3 \leqslant 1$.[6] *We have*

$$\begin{aligned}
&|\nabla_{\lambda,\lambda} L_i(\theta^1) - \nabla_{\lambda,\lambda} L_i(\theta^2)| \leqslant C(1+||\delta^1||_3)^7[1+|z^3_{i,N}(\lambda_1,\beta^1)| + |z^3_{i,N}(\lambda_2,\beta^2)|] \\
&\cdot \{[1+|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0} |x_{ik,N}|](w_{i\cdot,N}Y_N)^2 + 1\} \cdot ||\theta^1 - \theta^2||_2,
\end{aligned} \tag{D.10}$$

$$\begin{aligned}
&|\nabla_{\lambda,\beta_k} L_i(\theta^1) - \nabla_{\lambda,\beta_k} L_i(\theta^2)| \leqslant C(1+||\delta^1||_3)^7[1+|z^3_{i,N}(\lambda_1,\beta^1)| + |z^3_{i,N}(\lambda_2,\beta^2)|] \\
&\cdot [1+|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0} |x_{ik,N}|]|x_{ik,N} \cdot w_{i\cdot,N}Y_N| \cdot ||\theta^1 - \theta^2||_2,
\end{aligned} \tag{D.11}$$

$$\begin{aligned}
&|\nabla_{\beta_j,\beta_k} L_i(\theta^1) - \nabla_{\beta_j,\beta_k} L_i(\theta^2)| \leqslant C(1+||\delta^1||_3)^7[1+|z^3_{i,N}(\lambda_1,\beta^1)| + |z^3_{i,N}(\lambda_2,\beta^2)|] \\
&\cdot [1+|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0} |x_{ik,N}|]|x_{ik,N}x_{ij,N}| \cdot ||\theta^1 - \theta^2||_2,
\end{aligned} \tag{D.12}$$

$$\begin{aligned}
&|\nabla_{\lambda,\delta_k} L_{i,N}(\theta^1) - \nabla_{\lambda,\delta_k} L_{i,N}(\theta^2)| \leqslant Ck^2(1+||\delta^1||_2)^7 \cdot \\
&[1+z^2_{i,N}(\lambda_2,\beta^2) + z^2_{i,N}(\lambda_1,\beta^1)][1+|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0} |x_{ik,N}|]|w_{i\cdot,N}Y_N| \cdot ||\theta^1 - \theta^2||_1,
\end{aligned} \tag{D.13}$$

$$\begin{aligned}
&|\nabla_{\beta_j,\delta_k} L_{i,N}(\theta^1) - \nabla_{\beta_j,\delta_k} L_{i,N}(\theta^2)| \leqslant Ck^2(1+||\delta^1||_2)^7 \cdot \\
&[1+z^2_{i,N}(\lambda_2,\beta^2) + z^2_{i,N}(\lambda_1,\beta^1)][1+|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0} |x_{ik,N}|]|x_{ij,N}| \cdot ||\theta^1 - \theta^2||_1,
\end{aligned} \tag{D.14}$$

[6] The condition $||\theta^1 - \theta^2||_3 \leqslant 1$ is added to simplify the bounds. Because these results are used for $\theta^1 = \theta^0$ and $\theta^2 = \hat{\theta}_n$, by consistency, $||\hat{\theta}_n - \theta^0||_3 = o_p(1)$, so this restriction can be imposed without loss of generality.

$$
\begin{aligned}
&|\nabla_{\delta_j,\delta_k} L_i(\theta^1) - \nabla_{\delta_j,\delta_k} L_i(\theta^2)| \leqslant C(j+k)(1+||\delta^1||_1)^7 \cdot \\
&[1+|z_{i,N}(\lambda_1,\beta^1)|+|z_{i,N}(\lambda_2,\beta^2)|][1+|w_{i\cdot,N}Y_N|+\sum_{k=1}^{K^0}|x_{ik,N}|]\cdot||\theta^1-\theta^2||_0.
\end{aligned}
\tag{D.15}
$$

Proof of Lemma D.2: (1) Consider $\frac{1}{N}\nabla_{\lambda,\lambda}\ln L_N(\theta)$ first, which contains two parts from Eq. (D.4). By Assumption 13 and Lemmas B.2 - B.3, the first part is $\leqslant C_1(1+||\delta||_2^4)[1+z_{i,N}^2(\lambda,\beta)](w_{i\cdot,N}Y_N)^2$ for some constant $C_1>0$. And the second term is uniformly bounded (in λ):

$$
\sup_{\lambda\in\Lambda}|\sum_{k=2}^{\infty}(k-1)\lambda^{k-2}(\widetilde{W_N}^k)_{ii}| \leqslant \sum_{k=2}^{\infty}(k-1)\lambda_m^{k-2}||\widetilde{W_N}^k||_\infty \leqslant \lambda_m^{-2}\sum_{k=2}^{\infty}(k-1)\zeta_m^k<\infty.
$$

The other inequalities are similarly obtained.

(2) To estimate bounds of both terms on the right hand side, we will utilize Assumption 13 and Lemmas C.2 and B.1 - B.3 repeatedly. C_1, C_2, $\cdots$ are different positive numbers that do not depend on i, j, k N or M. C_i and C_j in the proof of different inequalities, e.g., the C_1 in the proof for Eq. (D.10) and the C_1 in that for Eq. (D.13), might be different.

For Eq. (D.10), we have $|\nabla_{\lambda,\lambda}L_i(\theta^1)-\frac{1}{N}\nabla_{\lambda,\lambda}L_i(\theta^2)| \leqslant |\nabla_{\lambda,\lambda}L_i(\lambda_1,\beta^1,\delta^1)-\nabla_{\lambda,\lambda}L_i(\lambda_2,\beta^2,\delta^1)|+|\nabla_{\lambda,\lambda}L_i(\lambda_2,\beta^2,\delta^1)-\nabla_{\lambda,\lambda}L_i(\lambda_2,\beta^2,\delta^2)|$. Consider $|\nabla_{\lambda,\lambda}L_i(\lambda_1,\beta^1,\delta^1)-\nabla_{\lambda,\lambda}L_i(\lambda_2,\beta^2,\delta^1)|$ first. By Assumption 13 and Lemmas B.1 - B.3, we have

$$
\begin{aligned}
&|\frac{d}{dz}[h(G(z)|\delta)\psi_1(G(z)|\delta)\frac{g'(z)}{G(z)}]| \\
=&|h'(G(z)|\delta)g(z)\psi_1(G(z)|\delta)\frac{g'(z)}{G(z)}+h(G(z)|\delta)\psi_1'(G(z)|\delta)G(z)\frac{g(z)}{G(z)}\frac{g'(z)}{G(z)} \\
&+h(G(z)|\delta)\psi_1(G(z)|\delta)[\frac{g''(z)}{G(z)}-\frac{g'(z)g(z)}{G^2(z)}]| \\
\leqslant&C_1\{(1+||\delta||_1)^2\cdot(1+|z|)^2+(1+||\delta||_0)^2\cdot(1+||\delta||_0)^2\cdot(1+|z|)\cdot(1+|z|)^2 \\
&+(1+||\delta||_0)^2\cdot[(1+|z|)^3+(1+|z|)^3]\}\leqslant 4C_1(1+||\delta||_1)^4(1+|z|)^3,
\end{aligned}
\tag{D.16}
$$

$$\begin{aligned}
&|\frac{d}{dz}[h'(G(z)|\delta)\psi_1(G(z)|\delta)\frac{g^2(z)}{G(z)}]| \\
=&|h''(G(z)|\delta)g(z)\psi_1(G(z)|\delta)\frac{g^2(z)}{G(z)} + h'(G(z)|\delta)\psi_1'(G(z)|\delta)g(z)\frac{g^2(z)}{G(z)} \\
&+ h'(G(z)|\delta)\psi_1(G(z)|\delta)[\frac{2g(z)g'(z)}{G(z)} - \frac{g^3(z)}{G^2(z)}]| \\
\leqslant& C_2\{(1+||\delta||_2)^2\cdot(1+|z|) + (1+||\delta||_1)^2\cdot(1+||\delta||_0)^2\cdot(1+|z|)^2 \\
&+ (1+||\delta||_1)^2\cdot[(1+|z|)^2+(1+|z|)^3]\} \leqslant 4C_2(1+||\delta||_2)^4(1+|z|)^3,
\end{aligned} \tag{D.17}$$

$$\begin{aligned}
&|\frac{d}{dz}[h(G(z)|\delta)\psi_1(G(z)|\delta)\frac{g(z)}{G(z)}]^2| \\
=&|2h(G(z)|\delta)\psi_1(G(z)|\delta)\frac{g(z)}{G(z)}\cdot\{h'(G(z)|\delta)g(z)\psi_1(G(z)|\delta)\frac{g(z)}{G(z)} + \\
&h(G(z)|\delta)\psi_1'(G(z)|\delta)g(z)\frac{g(z)}{G(z)} + h(G(z)|\delta)\psi_1(G(z)|\delta)[\frac{g'(z)}{G(z)} - \frac{g^2(z)}{G^2(z)}]| \\
\leqslant& C_3(1+||\delta||_0)^2(1+|z|)\cdot(1+||\delta||_1)^4(1+|z|)^2 = C_3(1+||\delta||_1)^6(1+|z|)^3,
\end{aligned} \tag{D.18}$$

$$\begin{aligned}
&|\frac{d}{dz}[\frac{h'(G(z)|\delta)}{h(G(z)|\delta)}g'(z)]| \\
=&\left|\frac{h''(G(z)|\delta)g(z)}{h(G(z)|\delta)}g(z)g'(z) + \frac{h'(G(z)|\delta)}{h(G(z)|\delta)}g''(z) - [\frac{h'(G(z)|\delta)}{h(G(z)|\delta)}]^2 g(z)g'(z)\right| \\
\leqslant& C_4\{(1+||\delta||_2)^2 + (1+||\delta||_1)^2 + (1+||\delta||_1)^4\} \leqslant 3C_4(1+||\delta||_2)^4,
\end{aligned} \tag{D.19}$$

$$\begin{aligned}
&|\frac{d}{dz}[\frac{h''(G(z)|\delta)}{h(G(z)|\delta)}g^2(z)]| \\
=&\left|\frac{h'''(G(z)|\delta)}{h(G(z)|\delta)}g^3(z) + \frac{h''(G(z)|\delta)}{h(G(z)|\delta)}2g(z)g'(z) - \frac{h''(G(z)|\delta)h'(G(z)|\delta)}{h^2(G(z)|\delta)}g^3(z)\right| \\
\leqslant& C_5\{(1+||\delta||_3)^2 + (1+||\delta||_2)^2 + (1+||\delta||_2)^2\cdot(1+||\delta||_1)^2\} = 3C_5(1+||\delta||_3)^4
\end{aligned} \tag{D.20}$$

$$|\frac{d}{dz}[\frac{h'(G(z)|\delta)}{h(G(z)|\delta)}g'(z)]^2| = \left|2\frac{h'(G(z)|\delta)}{h(G(z)|\delta)}g'(z)\cdot\frac{d}{dz}[\frac{h'(G(z)|\delta)}{h(G(z)|\delta)}g'(z)]\right| \leqslant C_6(1+||\delta||_2)^6, \tag{D.21}$$

$$\left|\frac{d}{dz}\left[\frac{g''(z)}{g(z)} - \frac{g'(z)^2}{g(z)^2}\right]\right| = \left|\frac{g'''(z)}{g(z)} - \frac{3g'(z)g''(z)}{g(z)^2} + \frac{2g'(z)^3}{g(z)^3}\right| \leqslant 6c^3(1+|z|)^3 \tag{D.22}$$

and $|\frac{d}{d\lambda}\sum_{k=1}^{\infty} k\lambda^{k-1}(\widetilde{W_N}^{k+1})_{ii}| = |\sum_{k=2}^{\infty} k(k-1)\lambda^{k-1}(\widetilde{W_N}^{k+1})_{ii}| \leqslant \lambda_m^{-2}\sum_{k=2}^{\infty} k(k-1)\zeta^{k+1} < \infty.$

Thus,

$$|\nabla_{\lambda,\lambda}L_i(\lambda_1,\beta^1,\delta^1)-\nabla_{\lambda,\lambda}L_i(\lambda_2,\beta^2,\delta^1)|\leqslant C_7(1+||\delta^1||_3)^6[1+|z^3_{i,N}(\lambda_1,\beta^1)|+$$
$$|z^3_{i,N}(\lambda_2,\beta^2)|]\cdot\left[(1+|w_{i\cdot,N}Y_N|)\cdot|\lambda_1-\lambda_2|+\sum_{k=1}^{K^0}|x_{ik,N}|\cdot|\beta_{1k}-\beta_{2k}|\right](w_{i\cdot,N}Y_N)^2. \tag{D.23}$$

Next, consider $|\nabla_{\lambda,\lambda}L_i(\lambda_2,\beta^2,\delta^1)-\nabla_{\lambda,\lambda}L_i(\lambda_2,\beta^2,\delta^2)|$. Denote $u=G(z_{i,N}(\lambda,\beta))$. With the condition $||\theta^1-\theta^2||_3\leqslant 1$, by Lemmas C.2 and B.1 - B.3,

$$\begin{aligned}
&\left|h(u|\delta^1)\psi_1(u|\delta^1)-h(u|\delta^2)\psi_1(u|\delta^2)\right|\cdot\frac{|g'(z_{i,N}(\lambda,\beta))|}{G(z_{i,N}(\lambda,\beta))}\\
\leqslant&C_8[(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0+(1+||\delta^1||_0)^2\cdot(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0](1+|z_{i,N}(\lambda,\beta)|)^2\\
\leqslant&2C_8(1+||\delta^1||_0)^5(1+|z_{i,N}(\theta)|)^2||\delta^1-\delta^2||_0,
\end{aligned}$$

$$\begin{aligned}
&\left|h'(u|\delta^1)\psi_1(u|\delta^1)-h'(u|\delta^2)\psi_1(u|\delta^2)\right|\cdot\frac{g^2(z_{i,N}(\lambda,\beta))}{G(z_{i,N}(\lambda,\beta))}\\
\leqslant&C_9[(1+||\delta^1||_1)^3||\delta^1-\delta^2||_1+(1+||\delta^1||_1)^2\cdot(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0](1+|z_{i,N}(\lambda,\beta)|)\\
\leqslant&2C_9(1+||\delta^1||_1)^5||\delta^1-\delta^2||_1(1+|z_{i,N}(\lambda,\beta)|),
\end{aligned}$$

$$\begin{aligned}
&\left|h^2(u|\delta^1)\psi_1^2(u|\delta^1)-h^2(u|\delta^2)\psi_1^2(u|\delta^2)\right|\cdot\frac{g^2(z_{i,N}(\lambda,\beta))}{G^2(z_{i,N}(\lambda,\beta))}\\
\leqslant&\left|h(u|\delta^1)\psi_1(u|\delta^1)+h(u|\delta^2)\psi_1(u|\delta^2)\right|\cdot\left|h(u|\delta^1)\psi_1(u|\delta^1)-h(u|\delta^2)\psi_1(u|\delta^2)\right|(1+|z_{i,N}(\lambda,\beta)|)^2\\
\leqslant&C_{10}(1+||\delta^1||_0)^7(1+|z_{i,N}(\lambda,\beta)|)^2||\delta^1-\delta^2||_0,
\end{aligned}$$

$$\begin{aligned}
&|\frac{h'(u|\delta^1)}{h(u|\delta^1)}g'(z_{i,N}(\lambda,\beta))-\frac{h'(u|\delta^2)}{h(u|\delta^2)}g'(z_{i,N}(\lambda,\beta))|\\
\leqslant&\left[\frac{|h'(u|\delta^1)-h'(u|\delta^2)|}{h(u|\delta^1)}+|h'(u|\delta^2)|\frac{|h(u|\delta^1)-h(u|\delta^2)|}{h(u|\delta^1)h(u|\delta^2)}\right]\cdot|g'(z_{i,N}(\lambda,\beta))|\\
\leqslant&C_{11}[(1+||\delta^1||_1)^3||\delta^1-\delta^2||_1+(1+||\delta^1||_1)^2\cdot(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0]\\
\leqslant&2C_{11}(1+||\delta^1||_1)^5||\delta^1-\delta^2||_1,
\end{aligned}$$

$$
\begin{aligned}
&|\frac{h''(u|\delta^1)}{h(u|\delta^1)}g^2(z_{i,N}(\lambda,\beta)) - \frac{h''(u|\delta^2)}{h(u|\delta^2)}g^2(z_{i,N}(\lambda,\beta))| \\
\leqslant &\left[\frac{|h''(u|\delta^1) - h''(u|\delta^2)|}{h(u|\delta^1)} + |h''(u|\delta^2)|\frac{|h(u|\delta^1) - h(u|\delta^2)|}{h(u|\delta^1)h(u|\delta^2)}\right] \cdot g^2(z_{i,N}(\lambda,\beta)) \\
\leqslant &C_{12}[(1+||\delta^1||_2)^3||\delta^1-\delta^2||_2 + (1+||\delta^1||_2)^2 \cdot (1+||\delta^1||_0)^3||\delta^1-\delta^2||_0] \\
\leqslant &2C_{12}(1+||\delta^1||_2)^5||\delta^1-\delta^2||_2,
\end{aligned}
$$

$$
\begin{aligned}
&\left|\left[\frac{h'(u|\delta^1)}{h(u|\delta^1)}\right]^2 - \left[\frac{h'(u|\delta^2)}{h(u|\delta^2)}\right]^2\right| g^2(z_{i,N}(\lambda,\beta)) \\
\leqslant &\left[\frac{h'(u|\delta^1)}{h(u|\delta^1)} + \frac{h'(u|\delta^2)}{h(u|\delta^2)}\right]\left[\frac{h'(u|\delta^1)}{h(u|\delta^1)} - \frac{h'(u|\delta^2)}{h(u|\delta^2)}\right] \cdot g^2(z_{i,N}(\lambda,\beta)) \\
\leqslant &C_{13}(1+||\delta^1||_1)^7||\delta^1-\delta^2||_1.
\end{aligned}
$$

Thus, $|\nabla_{\lambda,\lambda}L_i(\lambda_2,\beta^2,\delta^1) - \nabla_{\lambda,\lambda}L_i(\lambda_2,\beta^2,\delta^2)| \leqslant C_{14}(1+||\delta^1||_2)^7[1+|z_{i,N}(\lambda_2,\beta_2)|]^2(w_{i\cdot,N}Y_N)^2||\delta^1 - \delta^2||_2$. Together with Eq. (D.23),

$$
\begin{aligned}
&|\nabla_{\lambda,\lambda}L_i(\theta^1) - \nabla_{\lambda,\lambda}L_i(\theta^2)| \leqslant C_{15}(1+||\delta^1||_3)^7[1+|z^3_{i,N}(\lambda_1,\beta^1)| + |z^3_{i,N}(\lambda_2,\beta^2)|] \\
&\cdot \{[1+|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0}|x_{ik,N}|](w_{i\cdot,N}Y_N)^2 + 1\} \cdot ||\theta^1-\theta^2||_2.
\end{aligned}
$$

The proofs for Eq. (D.11) and (D.12) are similar to that for Eq. (D.10).

For Eq. (D.13), we have $|\nabla_{\lambda,\delta_k}L_i(\theta^1) - \nabla_{\lambda,\delta_k}L_i(\theta^2)| \leqslant |\nabla_{\lambda,\delta_k}L_i(\lambda_1,\beta^1,\delta^1) - \nabla_{\lambda,\delta_k}L_i(\lambda_2,\beta^2,\delta^1)| + |\nabla_{\lambda,\delta_k}L_i(\lambda_2,\beta^2,\delta^1) - \nabla_{\lambda,\delta_k}L_i(\lambda_2,\beta^2,\delta^2)|$. By Lemmas B.1 - B.3,

$$
\begin{aligned}
&|\frac{d}{dz}[\nabla_{\delta_k}h(G(z)|\delta)\psi_1(G(z)|\delta)\frac{g(z)}{G(z)}]| \\
=&|\nabla_{\delta_k}h'(G(z)|\delta)g(z)\psi_1(G(z)|\delta)\frac{g(z)}{G(z)} + \nabla_{\delta_k}h(G(z)|\delta)\psi_1'(G(z)|\delta)G(z)\frac{g(z)^2}{G(z)^2} \\
&+ \nabla_{\delta_k}h(G(z)|\delta)\psi_1(G(z)|\delta)[\frac{g'(z)}{G(z)} - \frac{g(z)^2}{G(z)^2}]| \\
\leqslant &C_1\{k(1+||\delta||_1)^2 \cdot (1+|z|) + (1+||\delta||_0)^2 \cdot (1+||\delta||_0)^2 \cdot (1+|z|)^2 + (1+||\delta||_0)^2 \cdot (1+|z|)^2\} \\
\leqslant &3C_1k(1+||\delta||_1)^4(1+|z|)^2,
\end{aligned}
\tag{D.24}
$$

$$
\begin{aligned}
&|\frac{d}{dz}[\frac{\nabla_{\delta_k}H(G(z)|\delta)}{G(z)}h(G(z)|\delta)\psi_1^2(G(z)|\delta)\frac{g(z)}{G(z)}]| \\
=&\Big|\Big[\Big(u\frac{d}{du}\frac{\nabla_{\delta_k}H(u|\delta)}{u}\Big)\,|_{u=G(z)}\ h(G(z)|\delta)\frac{g(z)^2}{G(z)^2}+\frac{\nabla_{\delta_k}H(G(z)|\delta)}{G(z)}h'(G(z)|\delta)\frac{g^2(z)}{G(z)}\Big]\psi_1^2(G(z)|\delta) \\
&+\frac{\nabla_{\delta_k}H(G(z)|\delta)}{G(z)}h(G(z)|\delta)\Big[2\psi_1(G(z)|\delta)\psi_1'(G(z)|\delta)\frac{g^2(z)}{G(z)}+\psi_1^2(G(z)|\delta)[\frac{g'(z)}{G(z)}-\frac{g(z)^2}{G(z)^2}]\Big]\Big| \\
\leqslant& C_2\{(1+||\delta||_0)^2\cdot(1+||\delta||_0)^2\cdot(1+|z|)^2+(1+||\delta||_0)^2\cdot(1+||\delta||_1)^2\cdot(1+|z|)+ \\
&(1+||\delta||_0)^2\cdot(1+||\delta||_0)^2\cdot[(1+||\delta||_0)^2\cdot(1+|z|)^2+(1+|z|)^2]\} \\
\leqslant& 4C_2(1+||\delta||_1)^6(1+|z|)^2,
\end{aligned}
\tag{D.25}
$$

$$
\begin{aligned}
&|\frac{d}{dz}[\frac{\nabla_{\delta_k}h'(G(z)|\delta)}{h(G(z)|\delta)}g(z)]| \\
=&\left|\frac{\nabla_{\delta_k}h''(G(z)|\delta)}{h(G(z)|\delta)}g(z)^2+\frac{\nabla_{\delta_k}h'(G(z)|\delta)}{h(G(z)|\delta)}g'(z)-\frac{\nabla_{\delta_k}h'(G(z)|\delta)}{h(G(z)|\delta)^2}h'(G(z)|\delta)g^2(z)\right| \\
\leqslant& C_3[k^2(1+||\delta||_2)^2+k(1+||\delta||_1)^2+k(1+||\delta||_1)^2\cdot(1+||\delta||_1)^2]\leqslant 3C_3k^2(1+||\delta||_2)^4
\end{aligned}
\tag{D.26}
$$

$$
\begin{aligned}
&|\frac{d}{dz}[\frac{\nabla_{\delta_k}h(G(z)|\delta)}{h^2(G(z)|\delta)}h'(G(z)|\delta)g(z)]| \\
=&\Big|\frac{\nabla_{\delta_k}h'(G(z)|\delta)}{h^2(G(z)|\delta)}h'(G(z)|\delta)g(z)^2+\frac{\nabla_{\delta_k}h(G(z)|\delta)}{h^2(G(z)|\delta)}h''(G(z)|\delta)g(z)^2+ \\
&\frac{\nabla_{\delta_k}h(G(z)|\delta)}{h^2(G(z)|\delta)}h'(G(z)|\delta)g'(z)-\frac{2\nabla_{\delta_k}h(G(z)|\delta)}{h^3(G(z)|\delta)}h'(G(z)|\delta)^2g^2(z)\Big| \\
\leqslant& C_4[k(1+||\delta||_1)^2\cdot(1+||\delta||_1)^2+(1+||\delta||_0)^2\cdot(1+||\delta||_2)^2+ \\
&(1+||\delta||_0)^2\cdot(1+||\delta||_1)^2+(1+||\delta||_0)^2\cdot(1+||\delta||_1)^4]\leqslant 4C_4k(1+||\delta||_2)^6.
\end{aligned}
\tag{D.27}
$$

Therefore,

$$
\begin{aligned}
&|\nabla_{\lambda,\delta_k}L_i(\lambda_1,\beta^1,\delta^1)-\nabla_{\lambda,\delta_k}L_i(\lambda_2,\beta^2,\delta^1)|\leqslant C_5k^2(1+||\delta^1||_2)^6\cdot \\
&[1+z_{i,N}^2(\lambda_2,\beta^2)+z_{i,N}^2(\lambda_1,\beta^1)][|w_{i\cdot,N}Y_N|\cdot|\lambda_1-\lambda_2|+\sum_{k=1}^{K^0}|x_{ik,N}|\cdot|\beta_{1k}-\beta_{2k}|].
\end{aligned}
\tag{D.28}
$$

Next, consider $|\nabla_{\lambda,\delta_k}L_i(\lambda_2,\beta^2,\delta^1)-\nabla_{\lambda,\delta_k}L_i(\lambda_2,\beta^2,\delta^2)|$. Denote $u=G(z_{i,N}(\lambda_2,\beta^2))$. With the condition $||\theta^1-\theta^2||_3 \leqslant 1$, by Lemmas C.2 and B.1 - B.3,

$$\begin{aligned}&\left|\nabla_{\delta_k} h(u|\delta^1)\psi_1(u|\delta^1) - \nabla_{\delta_k} h(u|\delta^2)\psi_1(u|\delta^2)\right| \frac{g(z_{i,N}(\lambda_2,\beta^2))}{G(z_{i,N}(\lambda_2,\beta^2))}\\ \leqslant & C_6[(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0 + (1+||\delta||_0)^2\cdot(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0](1+|z_{i,N}(\lambda_2,\beta^2)|)\\ \leqslant & 2C_6(1+||\delta^1||_0)^5(1+|z_{i,N}(\lambda_2,\beta^2)|)\cdot||\delta^1-\delta^2||_0,\end{aligned}$$

$$\begin{aligned}&\left|\frac{\nabla_{\delta_k} H(u|\delta^1)}{u} h(u|\delta^1)\psi_1^2(u|\delta^1) - \frac{\nabla_{\delta_k} H(u|\delta^2)}{u} h(u|\delta^2)\psi_1^2(u|\delta^2)\right| \frac{g(z_{i,N}(\lambda_2,\beta^2))}{G(z_{i,N}(\lambda_2,\beta^2))}\\ \leqslant & C_7[(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0 + (1+||\delta^1||_0)^2\cdot(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0](1+|z_{i,N}(\lambda_2,\beta^2)|)\\ \leqslant & 2C_7(1+||\delta^1||_0)^5(1+|z_{i,N}(\lambda_2,\beta^2)|)\cdot||\delta^1-\delta^2||_0,\end{aligned}$$

$$\begin{aligned}&\left|\frac{\nabla_{\delta_k} h'(u|\delta^1)}{h(u|\delta^1)} - \frac{\nabla_{\delta_k} h'(u|\delta^2)}{h(u|\delta^2)}\right| g(z_{i,N}(\lambda_2,\beta^2))\\ \leqslant & C_8[k(1+||\delta^1||_1)^3||\delta^1-\delta^2||_1 + k(1+||\delta^1||_1)^2\cdot(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0]\\ \leqslant & 2C_8 k(1+||\delta^1||_0)^5||\delta^1-\delta^2||_0,\end{aligned}$$

$$\begin{aligned}&\left|\frac{\nabla_{\delta_k} h(u|\delta^1)}{h^2(u|\delta^1)} h'(u|\delta^1) - \frac{\nabla_{\delta_k} h(u|\delta^2)}{h^2(u|\delta^2)} h'(u|\delta^2)\right| g(z_{i,N}(\lambda_2,\beta^2))\\ \leqslant & C_9[(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0\cdot(1+||\delta^1||_1)^2 + (1+||\delta^1||_0)^2\cdot(1+||\delta^1||_1)^3||\delta^1-\delta^2||_1 +\\ & (1+||\delta^1||_0)^2\cdot(1+||\delta^1||_1)^2\cdot(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0]\\ \leqslant & 3C_9(1+||\delta^1||_1)^7||\delta^1-\delta^2||_1.\end{aligned}$$

Thus,

$$|\nabla_{\lambda,\delta_k} L_i(\lambda_2,\beta^2,\delta^1) - \nabla_{\lambda,\delta_k} L_i(\lambda_2,\beta^2,\delta^2)| \leqslant C_{10}(1+||\delta^1||_1)^7[1+|z_{i,N}(\lambda_2,\beta^2)|]\cdot||\delta^1-\delta^2||_1. \quad \text{(D.29)}$$

By Eq. (D.28) and (D.29),

$$\begin{aligned}&|\nabla_{\lambda,\delta_k} L_{i,N}(\theta^1) - \nabla_{\lambda,\delta_k} L_{i,N}(\theta^2)| \leqslant C_{11} k^2 (1+||\delta^1||_2)^7\cdot\\ &[1+z_{i,N}^2(\lambda_2,\beta^2)+z_{i,N}^2(\lambda_1,\beta^1)][1+|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0}|x_{ik,N}|]|w_{i\cdot,N}Y_N|\cdot||\theta^1-\theta^2||_1.\end{aligned}$$

For Eq. (D.15), we have $|\nabla_{\delta_j,\delta_k} L_i(\theta^1) - \nabla_{\delta_j,\delta_k} L_i(\theta^2)| \leqslant |\nabla_{\delta_j,\delta_k} L_i(\lambda_1, \beta^1, \delta^1) - \nabla_{\delta_j,\delta_k} L_i(\lambda_2, \beta^2, \delta^1)| + |\nabla_{\delta_j,\delta_k} L_i(\lambda_2, \beta^2, \delta^1) - \nabla_{\delta_j,\delta_k} L_i(\lambda_2, \beta^2, \delta^2)|$. By Lemmas C.2 and B.1 - B.3,

$$\left|\frac{d}{dz}\frac{\nabla_{\delta_j,\delta_k} H(G(z)|\delta)}{H(G(z)|\delta)}\right| = \left|\left[u\frac{d}{du}\frac{\nabla_{\delta_j,\delta_k} H(u|\delta)}{H(u|\delta)}\right]\Big|_{u=G(z)} \cdot \frac{g(z)}{G(z)}\right| \leqslant C_1(1+||\delta||_0)^4(1+|z|), \quad \text{(D.30)}$$

$$\begin{aligned}
&\left|\frac{d}{dz}\left[\frac{\nabla_{\delta_j} H(G(z)|\delta)}{H(G(z)|\delta)}\frac{\nabla_{\delta_k} H(G(z)|\delta)}{H(G(z)|\delta)}\right]\right| \\
\leqslant &\left|\frac{\nabla_{\delta_j} h(G(z)|\delta)}{H(G(z)|\delta)} - \frac{\nabla_{\delta_j} H(G(z)|\delta) h(G(z)|\delta)}{H(G(z)|\delta)^2}\right| \frac{|\nabla_{\delta_k} H(G(z)|\delta)|}{H(G(z)|\delta)} g(z) + \\
&\frac{|\nabla_{\delta_j} H(G(z)|\delta)|}{H(G(z)|\delta)}\left|\frac{\nabla_{\delta_k} h(G(z)|\delta)}{H(G(z)|\delta)} - \frac{\nabla_{\delta_k} H(G(z)|\delta) h(G(z)|\delta)}{H(G(z)|\delta)^2}\right| g(z) \\
= &\left|\frac{\nabla_{\delta_j} h(G(z)|\delta)}{H(G(z)|\delta)/G(z)} - \frac{\frac{\nabla_{\delta_j} H(G(z)|\delta)}{G(z)} h(G(z)|\delta)}{[\frac{H(G(z)|\delta)}{G(z)}]^2}\right| \frac{|\nabla_{\delta_k} H(G(z)|\delta)|/G(z)}{H(G(z)|\delta)/G(z)}\frac{g(z)}{G(z)} + \qquad \text{(D.31)} \\
&\frac{|\nabla_{\delta_j} H(G(z)|\delta)/G(z)|}{H(G(z)|\delta)/G(z)}\left|\frac{\nabla_{\delta_k} h(G(z)|\delta)}{H(G(z)|\delta)/G(z)} - \frac{\frac{\nabla_{\delta_k} H(G(z)|\delta)}{G(z)} h(G(z)|\delta)}{[\frac{H(G(z)|\delta)}{G(z)}]^2}\right| \frac{g(z)}{G(z)} \\
\leqslant &C_2[(1+||\delta||_0)^2 + (1+||\delta||_0)^2 \cdot (1+||\delta||_0)^2](1+||\delta||_0)^2(1+|z|) \\
\leqslant &2C_2(1+||\delta||_0)^6(1+|z|),
\end{aligned}$$

$$\begin{aligned}
&\left|\frac{d}{dz}\frac{\nabla_{\delta_j,\delta_k} h(G(z)|\delta)}{h(G(z)|\delta)}\right| = \left|\frac{\nabla_{\delta_j,\delta_k} h'(G(z)|\delta)}{h(G(z)|\delta)} - \frac{\nabla_{\delta_j,\delta_k} h(G(z)|\delta)}{h(G(z)|\delta)^2} h'(G(z)|\delta)\right| g(z) \qquad \text{(D.32)} \\
\leqslant &C_3[(j+k)(1+||\delta||_1)^2 + (1+||\delta||_0)^2 \cdot (1+||\delta||_1)^2] \leqslant 2C_3(j+k)(1+||\delta||_1)^4,
\end{aligned}$$

$$\begin{aligned}
&\left|\frac{d}{dz}\frac{\nabla_{\delta_j} h(G(z)|\delta)\nabla_{\delta_k} h(G(z)|\delta)}{h^2(G(z)|\delta)}\right| = \left|\frac{\nabla_{\delta_j} h'(G(z)|\delta)\nabla_{\delta_k} h(G(z)|\delta)}{h^2(G(z)|\delta)} + \right. \\
&\left.\frac{\nabla_{\delta_j} h(G(z)|\delta)\nabla_{\delta_k} h'(G(z)|\delta)}{h^2(G(z)|\delta)} - \frac{\nabla_{\delta_j} h(G(z)|\delta)\nabla_{\delta_k} h(G(z)|\delta) h'(G(z)|\delta)}{h^3(G(z)|\delta)}\right| \cdot g(z) \qquad \text{(D.33)} \\
\leqslant &C_4[(1+||\delta||_1)^2 \cdot (1+||\delta||_0)^2 + (1+||\delta||_0)^2 \cdot (1+||\delta||_0)^2 \cdot (1+||\delta||_1)^2] \leqslant 2C_4(1+||\delta||_1)^6.
\end{aligned}$$

Thus,

$$|\nabla_{\delta_j,\delta_k} L_i(\lambda_1,\beta^1,\delta^1) - \nabla_{\delta_j,\delta_k} L_i(\lambda_2,\beta^2,\delta^1)| \leqslant C_5(j+k)(1+||\delta||_1)^6 \cdot$$
$$[1+|z_{i,N}(\lambda_1,\beta^1)| + |z_{i,N}(\lambda_2,\beta^2)|][|w_{i\cdot,N}Y_N|\cdot|\lambda_1-\lambda_2| + \sum_{k=1}^{K^0}|x_{ik,N}|\cdot|\beta_{1k}-\beta_{2k}|]. \tag{D.34}$$

Next, consider $|\nabla_{\delta_j,\delta_k} L_i(\lambda_2,\beta^2,\delta^1) - \nabla_{\delta_j,\delta_k} L_i(\lambda_2,\beta^2,\delta^2)|$. Denote $u = G(z_{i,N}(\lambda_2,\beta^2))$. With the condition $||\theta^1-\theta^2||_3 \leqslant 1$, by Lemmas C.2 and B.1 - B.3,

$$\begin{aligned}
&\left|\frac{\nabla_{\delta_j,\delta_k}H(u|\delta^1)}{H(u|\delta^1)} - \frac{\nabla_{\delta_j,\delta_k}H(u|\delta^2)}{H(u|\delta^2)}\right| \\
\leqslant & \frac{|\nabla_{\delta_j,\delta_k}H(u|\delta^1) - \nabla_{\delta_j,\delta_k}H(u|\delta^2)|/u}{H(u|\delta^1)/u} + \left|\frac{\nabla_{\delta_j,\delta_k}H(u|\delta^2)}{u}\right| \frac{|H(u|\delta^1)-H(u|\delta^2)|/u}{\frac{H(u|\delta^1)}{u}\frac{H(u|\delta^2)}{u}} \\
\leqslant & C_6[(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0 + (1+||\delta^1||_0)^2\cdot(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0] \\
\leqslant & 2C_6(1+||\delta^1||_0)^5||\delta^1-\delta^2||_0,
\end{aligned}$$

$$\begin{aligned}
&\left|\frac{\nabla_{\delta_j}H(u|\delta^1)}{H(u|\delta^1)}\frac{\nabla_{\delta_k}H(u|\delta^1)}{H(u|\delta^1)} - \frac{\nabla_{\delta_j}H(u|\delta^2)}{H(u|\delta^2)}\frac{\nabla_{\delta_k}H(u|\delta^2)}{H(u|\delta^2)}\right| \\
\leqslant & \left|\frac{\nabla_{\delta_j}H(u|\delta^1)}{H(u|\delta^1)} - \frac{\nabla_{\delta_j}H(u|\delta^2)}{H(u|\delta^2)}\right| \frac{|\nabla_{\delta_k}H(u|\delta^1)/u|}{H(u|\delta^1)/u} + \frac{|\nabla_{\delta_j}H(u|\delta^2)/u|}{H(u|\delta^2)/u}\left|\frac{\nabla_{\delta_k}H(u|\delta^1)}{H(u|\delta^1)} - \frac{\nabla_{\delta_k}H(u|\delta^2)}{H(u|\delta^2)}\right| \\
\leqslant & C_7(1+||\delta^1||_0)^2\left\{\left[\frac{|\nabla_{\delta_j}H(u|\delta^1) - \nabla_{\delta_j}H(u|\delta^2)|/u}{H(u|\delta^1)/u} + \left|\frac{\nabla_{\delta_j}H(u|\delta^2)}{u}\right|\frac{|H(u|\delta^1)-H(u|\delta^2)|/u}{\frac{H(u|\delta^1)}{u}\frac{H(u|\delta^2)}{u}}\right]\right. \\
&\left. + \left[\frac{|\nabla_{\delta_k}H(u|\delta^1) - \nabla_{\delta_k}H(u|\delta^2)|}{H(u|\delta^1)} + \left|\frac{\nabla_{\delta_k}H(u|\delta^2)}{u}\right|\frac{|H(u|\delta^1)-H(u|\delta^2)|/u}{\frac{H(u|\delta^1)}{u}\frac{H(u|\delta^2)}{u}}\right]\right\} \\
\leqslant & C_7(1+||\delta^1||_0)^2[(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0 + (1+||\delta^1||_0)^2\cdot(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0] \\
\leqslant & 2C_7(1+||\delta^1||_0)^7||\delta^1-\delta^2||_0,
\end{aligned}$$

$$\begin{aligned}
&\left|\frac{\nabla_{\delta_j,\delta_k}h(u|\delta^1)}{h(u|\delta^1)} - \frac{\nabla_{\delta_j,\delta_k}h(u|\delta^2)}{h(u|\delta^2)}\right| \\
\leqslant & \frac{|\nabla_{\delta_j,\delta_k}h(u|\delta^1) - \nabla_{\delta_j,\delta_k}h(u|\delta^2)|}{h(u|\delta^1)} + |\nabla_{\delta_j,\delta_k}h(u|\delta^2)|\cdot\frac{|h(u|\delta^1)-h(u|\delta^2)|}{h(u|\delta^1)h(u|\delta^2)} \\
\leqslant & C_8[(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0 + (1+||\delta^1||_0)^2\cdot(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0] \\
\leqslant & 2C_8(1+||\delta^1||_0)^5||\delta^1-\delta^2||_0,
\end{aligned}$$

$$
\begin{aligned}
&\left|\frac{\nabla_{\delta_j} h(u|\delta^1)\nabla_{\delta_k} h(u|\delta^1)}{h^2(u|\delta^1)} - \frac{\nabla_{\delta_j} h(u|\delta^2)\nabla_{\delta_k} h(u|\delta^2)}{h^2(u|\delta^2)}\right| \\
=&\left|\left[\frac{\nabla_{\delta_j} h(u|\delta^1)}{h(u|\delta^1)} - \frac{\nabla_{\delta_j} h(u|\delta^2)}{h(u|\delta^2)}\right]\frac{\nabla_{\delta_k} h(u|\delta^1)}{h(u|\delta^1)} + \frac{\nabla_{\delta_j} h(u|\delta^2)}{h(u|\delta^2)}\left[\frac{\nabla_{\delta_k} h(u|\delta^1)}{h(u|\delta^1)} - \frac{\nabla_{\delta_k} h(u|\delta^2)}{h(u|\delta^2)}\right]\right| \\
\leqslant&\left[\frac{|\nabla_{\delta_j} h(u|\delta^1) - \nabla_{\delta_j} h(u|\delta^2)|}{h(u|\delta^1)} + |\nabla_{\delta_j} h(u|\delta^2)|\frac{|h(u|\delta^1) - h(u|\delta^2)|}{h(u|\delta^1)h(u|\delta^2)}\right]\frac{|\nabla_{\delta_k} h(u|\delta^1)|}{h(u|\delta^1)} + \\
&\frac{|\nabla_{\delta_j} h(u|\delta^2)|}{h(u|\delta^2)}\left[\frac{|\nabla_{\delta_k} h(u|\delta^1) - \nabla_{\delta_k} h(u|\delta^2)|}{h(u|\delta^1)} + |\nabla_{\delta_k} h(u|\delta^2)|\frac{|h(u|\delta^2) - h(u|\delta^1)|}{h(u|\delta^1)h(u|\delta^2)}\right] \\
\leqslant& C_9[(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0 + (1+||\delta^1||_0)^2\cdot(1+||\delta^1||_0)^3||\delta^1-\delta^2||_0](1+||\delta^1||_0)^2 \\
\leqslant& 2C_9(1+||\delta^1||_0)^7||\delta^1-\delta^2||_0.
\end{aligned}
$$

Thus, $|\nabla_{\delta_j,\delta_k} L_i(\lambda_2,\beta^2,\delta^1) - \nabla_{\delta_j,\delta_k} L_i(\lambda_2,\beta^2,\delta^2)| \leqslant C_{10}(1+||\delta^1||_0)^7||\delta^1-\delta^2||_0$. With Eq. (D.34), we have

$$
\begin{aligned}
&|\nabla_{\delta_j,\delta_k} L_i(\theta^1) - \nabla_{\delta_j,\delta_k} L_i(\theta^2)| \leqslant C_{11}(j+k)(1+||\delta^1||_1)^7 \\
&[1+|z_{i,N}(\lambda_1,\beta^1)| + |z_{i,N}(\lambda_2,\beta^2)|][1+|w_{i\cdot,N}Y_N| + \sum_{k=1}^{K^0}|x_{ik,N}|]||\theta^1-\theta^2||_0.
\end{aligned} \tag{D.35}
$$

□

E. The Proof for Section 4–Asymptotic Distribution

The following lemma provides the covariance structure of a spatial NED process due to JP (2012).

Lemma E.1. (*Lemma A.3 in JP, 2012*) *Let Assumption 1 hold and let $X_{i,N}$ be uniformly L_2-NED on a random field $\{\epsilon_{i,N}\}$ with α-mixing coefficients $\alpha(u,v,s) \leqslant (u+v)^\tau \hat{\alpha}(s)$ for some $\tau > 0$: $||X_{i,N} - \mathrm{E}[X_{i,N}|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_X\psi(s)$, where C_X is a constant only depending on X and $\lim_{s\to\infty}\psi(s) = 0$. Let $S_N = \sum_{i=1}^N X_{i,N}$ and suppose that both $\hat{\alpha}(s)$ and $\psi(s)$ are nonincreasing. If $\sup_{i,N}||X_{i,N}||_{L^{2+\delta}} = A_X < \infty$ for some $\delta > 0$, then for any $i \neq j$, letting $d_{ij} = d(\vec{i},\vec{j})$, we have $|\operatorname{cov}(X_{i,N},\, X_{j,N})| \leqslant 4A_X^2[2C_d(\frac{d_{ij}}{3})^d]^{\tau_*}\hat{\alpha}^{\delta/(2+\delta)}(\frac{d_{ij}}{3}) + 2A_XC_X\psi(\frac{d_{ij}}{3})$, where $\tau_* = \tau\delta/(2+\delta)$ and C_d is a constant only depending on d. If in addition, if $\sum_{r=0}^\infty(r+1)^{d-1}\psi(r) < \infty$ and $\sum_{r=0}^\infty(r+1)^{d(1+\tau_*)-1}\hat{\alpha}^{\delta/(2+\delta)}(r) < \infty$, then for some constant $\overline{C_d}$ that depends only on d,*

$$
\operatorname{var}(S_N) \leqslant N[A_X^2 + 4A_X^2(2C_d)^{\tau_*}\overline{C_d}\sum_{r=0}^\infty(r+1)^{d(1+\tau_*)-1}\hat{\alpha}^{\delta/(2+\delta)}(r) + 2A_XC_X\overline{C_d}\sum_{r=0}^\infty(r+1)^{d-1}\psi(r)].
$$

Proof of Proposition 2: (1) The conclusions are deduced from Lemmas 3 and D.1.

(2) (i) Denote $Q_1(x|\delta) \equiv \frac{h(G(x)|\delta)g(x)}{H(G(x)|\delta)}$ and $Q_2(x|\delta) \equiv \frac{h'(G(x)|\delta)g(x)}{h(G(x)|\delta)}$. By Lemma B.1 and Assumption 13, there exists a constant $C_1 > 0$ such that

$$|Q_1'(x|\delta^0)| = \left| \frac{h'(G(x)|\delta^0)g^2(x)}{H(G(x)|\delta^0)} + \frac{h(G(x)|\delta^0)}{H(G(x)|\delta^0)/G(x)} \frac{g(x)}{G(x)} \frac{g'(x)}{g(x)} - \frac{h^2(G(x)|\delta^0)g^2(x)}{H^2(G(x)|\delta^0)} \right| \leqslant C_1(x^2+1).$$

As a result, by Lemma 4, $\{Q_1(z_{i,N}|\delta^0)\}_{i=1}^N$ is UG L_2-NED: for any $\gamma_1 \in (0, \frac{1}{2})$, there exists a constant $C_2 > 0$ such that $\left\|Q_1(z_{i,N}|\delta^0) - \mathrm{E}[Q_1(z_{i,N}|\delta^0)|\mathcal{F}_{i,N}(s)]\right\|_{L^2} \leqslant C_2\zeta^{\gamma_1 s/\bar{d}_0}$. Hence, by Lemma C.3, for any $\gamma_2 \in (0, \frac{1}{8})$, $\{1(y_{i,N} = 0)Q_1(z_{i,N}|\delta^0)w_{i\cdot,N}Y_N\}$ is UG L_2-NED with NED coefficient $\zeta^{\gamma_2 s/\bar{d}_0}$.

Again, by Lemma B.1 (1) and (7) and Assumption 13, there exists a constant $C_3 > 0$ such that

$$\left|Q_2'(x|\delta^0)\right| = \left| \frac{h''(G(x)|\delta^0)g^2(x)}{h(G(x)|\delta^0)} + \frac{h'(G(x)|\delta^0)g'(x)}{h(G(x)|\delta^0)} - [\frac{h'(G(x)|\delta^0)g(x)}{h(G(x)|\delta^0)}]^2 \right| \leqslant C_3,$$

Then, $\{Q_2(z_{i,N}|\delta^0)\}_{i=1}^N$ is UG L_2-NED: $||Q_2(z_{i,N}|\delta^0) - \mathrm{E}[Q_2(z_{i,N}|\delta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_3C_z\zeta^{s/\bar{d}_0}$. According to Lemma C.3, $\{1(y_{i,N} > 0)Q_2(z_{i,N}|\delta^0)w_{i\cdot,N}Y_N\}_{i=1}^N$ is UG L_2-NED with NED coefficient $\zeta^{\gamma_2 s/\bar{d}_0}$. In consequence, $\{\nabla_\lambda L_{i,N}(\theta^0)\}_{i=1}^N$ is UG L_2-NED with NED coefficient $\zeta^{\gamma_2 s/\bar{d}_0}$.

Similarly with only $x_{ik,N}$ in place of $w_{i,N}Y_n$ in the previous proof, for each $1 \leqslant k \leqslant K^0$, $\{\nabla_{\beta_k} L_{i,N}(\theta^0)\}_{i=1}^N$ is also UG L_2-NED with NED coefficient $\zeta^{\gamma_2 s/\bar{d}_0}$.

(ii) First, we study properties of $Q_{1k}(x|\delta^0) \equiv \nabla_{\delta_k} H(G(x)|\delta^0)/H(G(x)|\delta^0)$, the first term of Eq. (D.3). By Lemmas B.1 (1) and (2), B.2 (5) and Assumption 13, there exists a constant $C_4 > 0$ not depending on k,

$$|Q_{1k}'(x|\delta^0)| = \left| \frac{\nabla_{\delta_k} h(G(x)|\delta^0)}{H(G(x)|\delta^0)/G(x)} - \frac{\nabla_{\delta_k} H(G(x)|\delta^0)}{H(G(x)|\delta^0)} \frac{h(G(x)|\delta^0)}{H(G(x)|\delta^0)/G(x)} \right| \frac{g(x)}{G(x)} \leqslant C_4(|x|+1),$$

Thus, by Lemma 4, for any $\gamma_3 \in (0, \frac{1}{2})$, there exists a constant $C_5 > 0$ depending neither on i, N nor k, such that $||Q_{1k}(z_{i,N}|\delta^0) - \mathrm{E}[Q_{1k}(z_{i,N}|\delta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_5\zeta^{\gamma_3 s/\bar{d}_0}$. Accordingly, by Lemmas B.2 (5) and C.2, $\{1(y_{i,N} = 0)Q_{1k}(z_{i,N}|\delta^0)\}_{i=1}^N$ is also uniformly bounded and UG L_2-NED: for some constant $C_6 > 0$ that depends on neither i, N nor k

$$\begin{aligned} &||1(y_{i,N}=0)Q_{1k}(z_{i,N}|\delta^0) - \mathrm{E}[1(y_{i,N}=0)Q_{1k}(z_{i,N}|\delta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \\ \leqslant&||1(y_{i,N}=0)Q_{1k}(z_{i,N}|\delta^0) - \mathrm{E}[1(y_{i,N}=0)|\mathcal{F}_{i,N}(s)]\,\mathrm{E}[Q_{1k}(z_{i,N}|\delta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \\ \leqslant&[1+\epsilon_0^{-1}(2\sqrt{2}+4||\delta^0||_0)]||1(y_{i,N}=0) - \mathrm{E}[1(y_{i,N}=0)|\mathcal{F}_{i,N}(s)]||_{L^2}+ \\ &||Q_{1k}(z_{i,N}|\delta^0) - \mathrm{E}[Q_{1k}(z_{i,N}|\delta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_6\zeta^{s/3\bar{d}_0} \end{aligned} \tag{E.1}$$

Second, we examine the properties of $Q_{2k}(u|\delta^0) \equiv \nabla_{\delta_k} h(u|\delta^0)/h(u|\delta^0)$. By Lemma B.1, $|Q_{2k}(u|\delta)| \leqslant \epsilon_0^{-1}4(1+\sqrt{2}||\delta^0||_0)^2$ is bounded, and Lipschitz:

$$\begin{aligned} \sup_u |Q'_{2k}(u|\delta^0)| \leqslant& \sup_u \left[\left|\frac{\nabla_{\delta_k} h'(u|\delta^0)}{h(u|\delta^0)}\right| + \left|\frac{\nabla_{\delta_k} h(u|\delta^0)}{h(u|\delta^0)}\frac{h'(u|\delta^0)}{h(u|\delta^0)}\right|\right] \\ \leqslant&\epsilon_0^{-1}8(1+\sqrt{2}\pi||\delta^0||_1)^2\pi k + \epsilon_0^{-1}(1+\sqrt{2}||\delta^0||_0)^2\pi||\delta^0||_1(2/\epsilon_0)^{1/2} \leqslant C_7 k \end{aligned}$$

for some constant $C_7 > 0$. Therefore, $\{1(y_{i,N} > 0)Q_{2k}(G(z_{i,N})|\delta^0)\}_{i=1}^N$ is uniformly bounded (in i and N) and UG L_2-NED: there is a constant $C_8 > 0$ that depends on neither i, N, nor k, such that

$$\begin{aligned} &||1(y_{i,N}>0)Q_2(G(z_{i,N})|\delta^0) - \mathrm{E}[1(y_{i,N}>0)Q_2(G(z_{i,N})|\delta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \\ \leqslant&||1(y_{i,N}>0)Q_2(G(z_{i,N})|\delta^0) - \mathrm{E}[1(y_{i,N}>0)|\mathcal{F}_{i,N}(s)]\,\mathrm{E}[Q_2(G(z_{i,N})|\delta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \\ \leqslant&\epsilon_0^{-1}4(1+\sqrt{2}||\delta^0||_0)^2||1(y_{i,N}>0) - \mathrm{E}[1(y_{i,N}>0)|\mathcal{F}_{i,N}(s)]||_{L^2}+ \\ &||Q_2(G(z_{i,N})|\delta^0) - \mathrm{E}[Q_2(G(z_{i,N})|\delta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_8 k\zeta^{s/3\bar{d}_0}, \end{aligned} \tag{E.2}$$

where the second inequality comes from Lemma C.2. Hence, Eq. (D.3), (E.1) and (E.2) together imply $||\nabla_{\delta_k} L_{i,N}(\theta^0) - \mathrm{E}[\nabla_{\delta_k} L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant (C_8+C_6)k\zeta^{s/3\bar{d}_0}$. □

Proof of Lemma 6: By Eq. (10), $\mathrm{E}\sup_{u\in[0,1]}|V_n(u)| \leqslant \sqrt{2N}\sum_{k=1}^n 2^{-k}\sup_{1\leqslant i\leqslant N}\mathrm{E}\,|\nabla_k L_{i,N}(\theta^0) - \nabla_k L_{i,N}(\theta_n^0)|$. In the following, we examine $\sup_{1\leqslant i\leqslant N}\mathrm{E}\,|\nabla_k L_{i,N}(\theta^0) - \nabla_k L_{i,N}(\theta_n^0)|$ and their possible upper bounds. Recall $z_{i,N} \equiv z_{i,N}(\lambda_0, \beta_0)$.

(1) For $k=1$, we aim to bound $\mathrm{E}\,|\nabla_\lambda L_{i,N}(\theta^0) - \nabla_\lambda L_{i,N}(\theta_n^0)|$. There are two terms involving δ for this derivative in Eq. (D.1). The first term is $1(y_{i,N}=0)\frac{h(G(z_{i,N})|\delta^0)}{H(G(z_{i,N})|\delta^0)}g(z_{i,N})w_{i\cdot,N}Y_N$. By Lemmas B.1 (1), (3) and (7), and B.2 (1) and (2), there exist constants $C_1 > 0$ and $C_2 > 0$ such

that

$$
\begin{aligned}
&\sup_{1\leqslant i\leqslant N} \mathrm{E}\left|1(y_{i,N}=0)\left[\frac{h(G(z_{i,N})|\delta^0)}{H(G(z_{i,N})|\delta^0)}-\frac{h(G(z_{i,N})|\delta_n^0)}{H(G(z_{i,N})|\delta_n^0)}\right]g(z_{i,N})w_{i\cdot,N}Y_N\right| \\
\leqslant &\sup_{1\leqslant i\leqslant N} \mathrm{E}[\frac{|h(G(z_{i,N})|\delta^0)-h(G(z_{i,N})|\delta_n^0)|}{H(G(z_{i,N})|\delta^0)/G(z_{i,N})}\frac{g(z_{i,N})}{G(z_{i,N})}|w_{i\cdot,N}Y_N| \\
&+h(G(z_{i,N})|\delta_n^0)\frac{|H(G(z_{i,N})|\delta_n^0)-H(G(z_{i,N})|\delta^0)|}{H(G(z_{i,N})|\delta^0)H(G(z_{i,N})|\delta_n^0)}g(z_{i,N})|w_{i\cdot,N}Y_N|] \\
\leqslant &\sup_{1\leqslant i\leqslant N} \mathrm{E}\left\{C_1||\delta_n^0-\delta^0||_0\left[||\delta^0||_0^3+||\delta^0||_0^2\cdot\frac{||\delta^0||_0^3G(z_{i,N})^2}{H(G(z_{i,N})|\delta^0)H(G(z_{i,N})|\delta_n^0)}\right]\frac{g(z_{i,N})}{G(z_{i,N})}|w_{i\cdot,N}Y_N|\right\} \\
\leqslant &C_2||\delta_n^0-\delta^0||_0\cdot||\delta^0||_0^5.
\end{aligned}
\tag{E.3}
$$

For the second term $1(y_{i,N}>0)\frac{h'(G(z_{i,N})|\delta^0)}{h(G(z_{i,N})|\delta^0)}g(z_{i,N})w_{i\cdot,N}Y_N$ in $\nabla_\lambda L_{i,N}(\theta^0)$, by Lemma B.1 (3) and (7), there exist some constants C_3 and C_4 such that

$$
\begin{aligned}
&\sup_{1\leqslant i\leqslant N} \mathrm{E}\left|1(y_{i,N}>0)\left[\frac{h'(G(z_{i,N})|\delta^0)}{h(G(z_{i,N})|\delta^0)}-\frac{h'(G(z_{i,N})|\delta_n^0)}{h(G(z_{i,N})|\delta_n^0)}\right]g(z_{i,N})w_{i\cdot,N}Y_N\right| \\
\leqslant &\sup_{1\leqslant i\leqslant N} \mathrm{E}\left\{\frac{|h'(G(z_{i,N})|\delta^0)-h'(G(z_{i,N})|\delta_n^0)|}{h(G(z_{i,N})|\delta^0)}g(z_{i,N})|w_{i\cdot,N}Y_N|+\right. \\
&\left.\frac{|h'(G(z_{i,N})|\delta_n^0)|}{h(G(z_{i,N})|\delta_n^0)}\cdot\frac{|h(G(z_{i,N})|\delta^0)-h(G(z_{i,N})|\delta_n^0)|}{h(G(z_{i,N})|\delta^0)}g(z_{i,N})|w_{i\cdot,N}Y_N|\right\} \\
\leqslant &\sup_{1\leqslant i\leqslant N} \mathrm{E}\,C_3\{[||\delta^0||_0^3\cdot||\delta_n^0-\delta^0||_1|w_{i\cdot,N}Y_N|+||\delta^0||_0\cdot||\delta^0||_0^3||\delta_n^0-\delta^0||_0|w_{i\cdot,N}Y_N|]\} \\
\leqslant &C_4||\delta^0||_0^4\cdot||\delta_n^0-\delta^0||_1.
\end{aligned}
$$

Thus, $\sup_{1\leqslant i\leqslant N}\mathrm{E}\,|\nabla_\lambda L_{i,N}(\theta^0)-\nabla_\lambda L_{i,N}(\theta_n^0)|\leqslant\max(C_2,C_4)||\delta^0||_0^5\cdot||\delta^0-\delta_n^0||_1$.

Similarly, from Eq. (D.2), $\mathrm{E}\sup_{1\leqslant i\leqslant N}|\nabla_{\beta_k}L_{i,N}(\theta^0)-\nabla_{\beta_k}L_{i,N}(\theta_n^0)|\leqslant C_5||\delta^0||_0^5\cdot||\delta^0-\delta_n^0||_1$ for some constant $C_5>0$, for all $1\leqslant k\leqslant K^0$.

(2) Consider $|\nabla_{\delta_k}L_{i,N}(\theta^0)-\nabla_{\delta_k}L_{i,N}(\theta_n^0)|$. By Eq. (E.3), Lemmas B.1 (8) and B.2 (1) - (5),

there exists a constant $C_6 > 0$,

$$
\begin{aligned}
&|\nabla_{\delta_k} L_{i,N}(\theta^0) - \nabla_{\delta_k} L_{i,N}(\theta^0_n)| \\
\leqslant& \left[\left| \frac{\nabla_{\delta_k} H(u|\delta^0)}{u} \frac{u}{H(u|\delta^0)} - \frac{\nabla_{\delta_k} H(u|\delta^0_n)}{u} \frac{u}{H(u|\delta^0_n)} \right| + \left| \frac{\nabla_{\delta_k} h(u|\delta^0)}{h(u|\delta^0)} - \frac{\nabla_{\delta_k} h(u|\delta^0_n)}{h(u|\delta^0_n)} \right| \right] \Bigg|_{u=G(z_{i,N})} \\
\leqslant& \frac{|\nabla_{\delta_k} H(G(z_{i,N})|\delta^0) - \nabla_{\delta_k} H(G(z_{i,N})|\delta^0_n)|}{\epsilon_0 G(z_{i,N})} + 4(1+\sqrt{2}||\delta^0||_0)^2 \left| \frac{G(z_{i,N})}{H(G(z_{i,N})|\delta^0)} - \frac{G(z_{i,N})}{H(G(z_{i,N})|\delta^0_n)} \right| \\
&+ \frac{|\nabla_{\delta_k} h(G(z_{i,N})|\delta^0) - \nabla_{\delta_k} h(G(z_{i,N})|\delta^0_n)|}{h(G(z_{i,N})|\delta^0)} + 4(1+\sqrt{2}||\delta^0||_0)^2 \left| \frac{1}{h(G(z_{i,N})|\delta^0)} - \frac{1}{h(G(z_{i,N})|\delta^0_n)} \right| \\
\leqslant& ||\delta^0 - \delta^0_n||_0 \{ \epsilon_0^{-1} [(8||\delta^0||_0 + 4)(\sqrt{2} + ||\delta^0||_0)^2 + 1] + 4(1+\sqrt{2}||\delta^0||_0)^2 \epsilon_0^{-2} \cdot \\
&[2||\delta^0||_0 (1+\sqrt{2}||\delta^0||_0)^2 + 2\sqrt{2} + 4||\delta^0||_0] + (\epsilon_0^{-2} + \epsilon_0^{-1}) 4(1+\sqrt{2}||\delta^0||_0)^2 \cdot [2||\delta^0||_0(1+\sqrt{2}||\delta^0||_0)^2 + \\
&2\sqrt{2} + 4||\delta^0||_0] \} \leqslant C_6 ||\delta^0||_0^5 \cdot ||\delta^0 - \delta^0_n||_0.
\end{aligned}
$$

Let $C_7 \equiv \max(C_2, C_4, C_5, C_6)$. Then, $\mathrm{E}\sup_{u\in[0,1]} |V_n(u)| \leqslant C_7 ||\delta^0||_0^5 \cdot ||\delta^0 - \delta^0_n||_1 \cdot \sqrt{2N} \sum_{k=1}^n 2^{-k} = \sqrt{2} C_7 ||\delta^0||_0^5 \cdot ||\delta^0 - \delta^0_n||_1 \sqrt{N}$. Since $||\delta^0 - \delta^0_n||_1 \leqslant (n - K^0)^{-(l_0-1)} \sum_{m=n-K^0}^\infty m^{l_0} |\delta_{0m}| = o(n^{-(l_0-1)})$, $\mathrm{E}\sup_{u\in[0,1]} |V_n(u)| = o\left(n^{-(l_0-1)}\sqrt{N}\right) = o(1)$ under Assumption 16. □

Proof of Lemma 7: $\mathrm{E}\sup_{0\leqslant u\leqslant 1} |\widehat{Z_n}(u) - \widetilde{Z_N}(u)| \leqslant \sqrt{2} \sum_{k=n+1}^\infty 2^{-k} \mathrm{E} |\frac{1}{\sqrt{N}} \sum_{i=1}^N \nabla_k L_{i,N}(\theta^0)| \leqslant \sqrt{2} \sum_{k=n+1}^\infty 2^{-k} || \frac{1}{\sqrt{N}} \sum_{i=1}^N \nabla_k L_{i,N}(\theta^0)||_{L^2}$. By Proposition 2 and Lemma E.1, under Assumption 15, $||\frac{1}{\sqrt{N}} \sum_{i=1}^N \nabla_k L_{i,N}(\theta^0)||_{L^2} \leqslant C\sqrt{k - (K^0+1)} \leqslant C\sqrt{k}$ for some $C > 0$ depending on neither N nor k. Then, the conclusion holds because $\mathrm{E}[\sup_{0\leqslant u\leqslant 1} |\widehat{Z_n}(u) - \widetilde{Z_N}(u)|] \leqslant C\sqrt{2} \sum_{k=n+1}^\infty 2^{-k}\sqrt{k}$, which converges to zero as $N \to \infty$. □

Proof of Lemma 8: From Theorem 7.1 in Billingsley (1999), to show $\widetilde{Z_N} \Rightarrow Z$, it is sufficient to show (1) that finite-dimensional distributions of $\widetilde{Z_N}$ converge weakly to those of Z, and (2) $\{\widetilde{Z_N}\}_{N=K^0+2}^\infty$ is tight.

(1) By Proposition 2, $||\nabla_k L_{i,N}(\theta^0) - \mathrm{E}[\nabla_k L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_1 k \zeta^{\gamma s/\bar{d}_0}$ for some $\gamma \in (0, \frac{1}{8})$ and $C_1 > 0$ depending on neither i, k nor N. Then, $\{\hat{Z}_{i,N}(u) \equiv \sum_{k=1}^\infty [N^{-1/2} \nabla_k L_{i,N}(\theta^0)] \eta_k(u)\}_{i=1}^N$ is UG L_2-NED as

$$
\begin{aligned}
&||\widetilde{Z_N}(u) - \mathrm{E}[\widetilde{Z_N}(u)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant \sum_{k=1}^\infty 2^{-k}\sqrt{2} ||\nabla_k L_{i,N}(\theta^0) - \mathrm{E}[\nabla_k L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \\
\leqslant& C_1 \sum_{k=1}^\infty 2^{-k} k \zeta^{\gamma s/\bar{d}_0} = 2C_1 \zeta^{\gamma s/\bar{d}_0}.
\end{aligned}
$$

Then, with Assumptions 15 and 17, the CLT for NED random field in JP (2012) is applicable for the convergence of finite-dimensional distributions to normal distributions.[7]

(2) We apply Theorem 7.3 in Billingsley (1999) to show the tightness of $\{\widetilde{Z_N}\}_{N=2+K^0}^{\infty}$. The first condition in Theorem 7.3 holds because of the weak convergence of finite-dimensional distributions shown above. For the remaining second condition, it is sufficient to show that, for any $\epsilon > 0$, $\sup_{u_1,u_2\in[0,1],|u_1-u_2|<\epsilon} |\widetilde{Z_N}(u_1) - \widetilde{Z_N}(u_2)| = \epsilon \cdot O_p(1)$. By Proposition 2 and Lemma E.1, $||\frac{1}{\sqrt{N}}\sum_{i=1}^N \nabla_k L_{i,N}(\theta^0)||_{L^2} = \{\text{var}[\frac{1}{\sqrt{N}}\sum_{i=1}^N \nabla_k L_{i,N}(\theta^0)]\}^{1/2} \leqslant C_2 k$ for some constant $C_2 > 0$. Then, by the mean value theorem,

$$\begin{aligned}
&\mathrm{E} \sup_{u_1,u_2\in[0,1],|u_1-u_2|<\epsilon} |\widetilde{Z_N}(u_1) - \widetilde{Z_N}(u_2)| \leqslant \epsilon\sqrt{2}\pi \sum_{k=1}^{\infty} 2^{-k} k\, \mathrm{E}\, |\frac{1}{\sqrt{N}}\sum_{i=1}^N \nabla_k L_{i,N}(\theta^0)| \\
\leqslant &\epsilon\sqrt{2}\pi \sum_{k=1}^{\infty} 2^{-k} k ||\frac{1}{\sqrt{N}}\sum_{i=1}^N \nabla_k L_{i,N}(\theta^0)||_{L^2} \leqslant C_2\epsilon\sqrt{2}\pi \sum_{k=1}^{\infty} 2^{-k} k^{1.5} = \epsilon O_p(1).
\end{aligned}$$

Next, we show $\sup_{0\leqslant u_1,u_2\leqslant 1} |\Gamma(u_1,u_2)| < \infty$ as follows:

$$\begin{aligned}
&\sup_{0\leqslant u_1,u_2\leqslant 1} |\Gamma(u_1,u_2)| = \sup_{0\leqslant u_1,u_2\leqslant 1} \lim_{N\to\infty} \\
&\left|\mathrm{E}\left\{\sum_{k=1}^{\infty}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^N \nabla_k L_{i,N}(\theta^0)\right] 2^{-k}\sqrt{2}\cos k\pi u_1 \cdot \sum_{m=1}^{\infty}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^N \nabla_m L_{i,N}(\theta^0)\right] 2^{-m}\sqrt{2}\cos m\pi u_2\right\}\right| \\
\leqslant &2\limsup_{N\to\infty}\sum_{k=1}^{\infty}\sum_{m=1}^{\infty} 2^{-k-m}\left|\mathrm{E}\left\{\left[\frac{1}{\sqrt{N}}\sum_{i=1}^N \nabla_k L_{i,N}(\theta^0)\right]\left[\frac{1}{\sqrt{N}}\sum_{i=1}^N \nabla_m L_{i,N}(\theta^0)\right]\right\}\right| \\
\leqslant &2\limsup_{N\to\infty}\sum_{k=1}^{\infty}\sum_{m=1}^{\infty} 2^{-k-m}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N \nabla_k L_{i,N}(\theta^0)\right\|_{L^2}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N \nabla_m L_{i,N}(\theta^0)\right\|_{L^2} \\
= &2\limsup_{N\to\infty}\left[\sum_{k=1}^{\infty} 2^{-k}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N \nabla_k L_{i,N}(\theta^0)\right\|_{L^2}\right]^2 \leqslant 2C_2^2 \limsup_{N\to\infty}\left(\sum_{k=1}^{\infty} 2^{-k} k^{1/2}\right)^2 < \infty.
\end{aligned}$$

□

[7] We note that the condition $\liminf_{N\to\infty} N^{-1}\lambda_{\min}[\text{var}(\widetilde{Z_N}(u_1),\cdots,\widetilde{Z_N}(u_J))] > 0$ in JP (2012), which rules out singularity of limiting variance matrix is not needed for our lemma. Non-singular variance matrices are not necessarily for the functional central limit theorem in this lemma. So such minimum eigenvalue condition is not imposed.

Proof of Proposition 3: (1) Denote $\nabla_{\lambda,\lambda}L_{i,N}(\theta^0) \equiv [1(y_{i,N} = 0)\Upsilon_1(z_{i,N}) + 1(y_{i,N} > 0)\Upsilon_2(z_{i,N})](w_{i\cdot,N}Y_N)^2 - [((I_N - \lambda_0\widetilde{W_N})^{-1}\widetilde{W_N})]_{ii}^2$, where

$$\Upsilon_1(z) \equiv h(G(z)|\delta^0)\psi_1(G(z)|\delta^0)\frac{g'(z)}{G(z)} + h'(G(z)|\delta^0)\psi_1(G(z)|\delta^0)\frac{g^2(z)}{G(z)} - [\frac{h(G(z)|\delta^0)\psi_1(G(z)|\delta^0)g(z)}{G(z)}]^2,$$

$$\Upsilon_2(z) \equiv \frac{h'(G(z)|\delta^0)g'(z)}{h(G(z)|\delta^0)} + \frac{h''(G(z)|\delta^0)g^2(z)}{h(G(z)|\delta^0)} - [\frac{h'(G(z)|\delta^0)g(z)}{h(G(z)|\delta^0)}]^2 + \frac{g''(z)}{g(z)} - \frac{g'(z)^2}{g(z)^2}.$$

From the proof of Lemma D.2 (1), $\{\Upsilon_1(z_{i,N})\}_{i=1}^N$ and $\{\Upsilon_2(z_{i,N})\}_{i=1}^N$ are uniformly L_p bounded for any $p > 1$. With Lemma 3 and Eq. (D.16) - (D.22), by Lemma 4, for any $\gamma_1 \in (0, \frac{1}{2})$, $\{\Upsilon_1(z_{i,N})\}_{i=1}^N$ and $\{\Upsilon_2(z_{i,N})\}_{i=1}^N$ are UG L_2-NED with NED coefficient $\zeta^{\gamma_1 s/\bar{d}_0}$. Because $(w_{i\cdot,N}Y_N)^2$ is also UG L_2-NED with coefficients $\zeta^{\gamma_1 s/\bar{d}_0}$, by Lemma C.3, $\{\Upsilon_1(z_{i,N})(w_{i\cdot,N}Y_N)^2\}_{i=1}^N$ and $\{\Upsilon_2(z_{i,N})(w_{i\cdot,N}Y_N)^2\}_{i=1}^N$ are both UG L_2-NED with coefficients $\zeta^{\gamma_2 s/\bar{d}_0}$ for any $\gamma_2 \in (0, \frac{1}{4})$. From Proposition 3 (1) in XL (2015b), $\{[((I_N - \lambda_0\widetilde{W_N})^{-1}\widetilde{W_N})]_{ii}^2\}_{i=1}^N$ is UG L_2-NED with NED coefficient $s^2\zeta^{s/3\bar{d}_0}$. As a result, by Lemma C.3, for every $\gamma_3 \in (0, \frac{1}{8})$, there exists a constant C_1 such that $||\nabla_{\lambda,\lambda}L_{i,N}(\theta^0) - \mathrm{E}[\nabla_{\lambda,\lambda}L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_1\zeta^{\gamma_3 s/\bar{d}_0}$.

Similarly, $\{\nabla_{\lambda,\beta_k}L_{i,N}(\theta^0)\}$ and $\{\nabla_{\beta_j,\beta_k}L_{i,N}(\theta^0)\}$ are UG L_2-NED .

(2) $\nabla_{\lambda,\delta_k}L_{i,N}(\theta^0) = -[1(y_{i,N} = 0)\Upsilon_{3k}(z_{i,N}) + 1(y_{i,N} > 0)\Upsilon_{4k}(z_{i,N})]w_{i\cdot,N}Y_N$, where

$$\Upsilon_{3k}(z) \equiv [\nabla_{\delta_k}h(G(z)|\delta^0)\psi_1(G(z)|\delta^0) - \frac{h(G(z)|\delta^0)\psi_1^2(u|\delta^0)\nabla_{\delta_k}H(G(z)|\delta^0)}{G(z)}]\frac{g(z)}{G(z)},$$

$$\Upsilon_{4k}(z) \equiv [\frac{\nabla_{\delta_k}h'(G(z)|\delta^0)}{h(G(z)|\delta^0)} - \frac{\nabla_{\delta_k}h(G(z)|\delta^0)h'(G(z)|\delta^0)}{h^2(G(z)|\delta^0)}]g(z).$$

From the proof of Lemma D.2 (1), $\{\Upsilon_{3k}(z_{i,N})/k\}_{i=1}^N$ and $\{\Upsilon_{4k}(z_{i,N})/k\}_{i=1}^N$ are uniformly (in i, N and k) L_p bounded for any $p > 1$. With Lemma 3, Eq. (D.24) - (D.27), by Lemma 4, for any $\gamma_4 \in (0, \frac{1}{2})$, $\{\Upsilon_{3k}(z_{i,N})/k^2\}_{i=1}^N$ and $\{\Upsilon_{4k}(z_{i,N})/k^2\}_{i=1}^N$ are UG L_2-NED with NED coefficient $\zeta^{\gamma_4 s/\bar{d}_0}$. Because $\{w_{i\cdot,N}Y_N\}_{i=1}^N$ is UG L_2-NED with coefficients $\zeta^{s/\bar{d}_0}$, by Lemma C.3, $\gamma_5 \in (0, \frac{1}{4})$, $\{\Upsilon_{3k}(z_{i,N})w_{i\cdot,N}Y_N/k^2\}_{i=1}^N$ and $\{\Upsilon_{4k}(z_{i,N})w_{i\cdot,N}Y_N/k^2\}_{i=1}^N$ are both UG L_2-NED in i, N and k with coefficients $\zeta^{\gamma_5 s/\bar{d}_0}$. Accordingly, by Lemma C.3, for any $\gamma_6 \in (0, \frac{1}{8})$, there exist a constant C_2 such that $||\nabla_{\lambda,\delta_k}L_{i,N}(\theta^0) - \mathrm{E}[\nabla_{\lambda,\delta_k}L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_2k^2\zeta^{\gamma_6 s/\bar{d}_0}$.

The proof for the NED of $\{\nabla_{\beta_j,\delta_k}L_{i,N}(\theta^0)\}$ is similar.

(3) $\nabla_{\delta_j,\delta_k} L_{i,N}(\theta^0) = 1(y_{i,N} = 0)\Upsilon_{5jk}(z_{i,N}) + 1(y_{i,N} > 0)\Upsilon_{6jk}(z_{i,N})$, where

$$\Upsilon_{5jk}(z) \equiv \frac{\nabla_{\delta_j,\delta_k} H(G(z)|\delta^0)}{H(G(z)|\delta^0)} - \frac{\nabla_{\delta_j} H(G(z)|\delta^0)}{H(G(z)|\delta^0)} \frac{\nabla_{\delta_k} H(G(z)|\delta^0)}{H(G(z)|\delta^0)},$$

$$\Upsilon_{6jk}(z) \equiv \frac{\nabla_{\delta_j,\delta_k} h(G(z)|\delta^0)}{h(G(z)|\delta^0)} - \frac{\nabla_{\delta_j} h(G(z)|\delta^0) \nabla_{\delta_k} h(G(z)|\delta^0)}{h^2(G(z)|\delta^0)}.$$

By Lemmas B.1 - B.2, all terms in $\Upsilon_{5jk}(z)$ and $\Upsilon_{6jk}(z)$ are uniformly (in j and k) bounded. By Eq. (D.30) - (D.33) and Lemma 4, for any $\gamma_7 \in (0, \frac{1}{2})$, $\{\Upsilon_{5k}(z_{i,N})/(j+k)\}_{i=1}^N$ and $\{\Upsilon_{6k}(z_{i,N})/(j+k)\}_{i=1}^N$ are UG L_2-NED in i, N, j and k, with NED coefficient $\zeta^{\gamma_7 s/\bar{d}_0}$. Because of uniform boundedness of $\Upsilon_{5jk}(z)$ and $\Upsilon_{6jk}(z)$, $||\nabla_{\delta_j,\delta_k} L_{i,N}(\theta^0) - \mathrm{E}[\nabla_{\delta_j,\delta_k} L_{i,N}(\theta^0)|\mathcal{F}_{i,N}(s)]||_{L^2} \leqslant C_3(j+k)\zeta^{s/3\bar{d}_0}$ for some constant $C_3 > 0$ that depends on neither i, N, j nor k. $\square$

The proof of Lemma **9** is built on Theorem B.1 in Bierens (2014):

Lemma E.2. *Let Y_N and $X_{1,N}$, $X_{2,N}$, $\cdots$, $X_{n,N}$ be random elements*[8] *of a Hilbert space $(\mathcal{H}, ||\cdot||)$, on the basis on a sample of size N, where n is a subsequence of N. Let $\hat{Y}_{n,N}$ be the projection of Y_N on $\mathrm{span}(\{X_{m,N}\}_{m=1}^n)$, with residual $U_{n,N} = Y_N - \hat{Y}_{n,N}$. Suppose that the following conditions holds.*

(a) There exists a nonrandom element $y \in \mathcal{H}$ such that $\mathrm{plim}_{N\to\infty} ||Y_N - y|| = 0$.

(b) There exists a sequence $\{x_m\}_{m=1}^\infty$ of nonrandom elements of $\mathcal{H}$ and a sequence $\{\rho_m\}_{m=1}^\infty$ of positive numbers such that $\mathrm{plim}_{N\to\infty} \sum_{m=1}^n \rho_m ||X_{m,N} - x_m|| = 0$.

(c) $\liminf_{n\to\infty} ||\sum_{m=1}^n \rho_m x_m|| > 0$.

Then, $\mathrm{plim}_{N\to\infty} ||\hat{Y}_{n,N} - \hat{y}|| = 0$ and $\mathrm{plim}_{N\to\infty} ||U_{n,N} - u|| = 0$, where $\hat{y}$ is the projection of y on $\mathrm{span}(\{x_m\}_{m=1}^\infty)$ and $u = y - \hat{y}$ is the residual.

Proof of Lemma 9: For any function $c(u) \in L^2(0,1)$, define $||c(\cdot)|| \equiv [\int_0^1 c^2(u)du]^{1/2}$. To have the conclusion, we shall check conditions (a) - (c) in Lemma E.2. For conditions (a) and (b), it suffices to show that there exists a summable positive sequence $\{\rho_m\}_{m=1}^\infty$ such that $\sum_{m=1}^n \rho_m ||\hat{b}_{m,n} - b_m||^2 = o_p(1)$. We can let ρ_m be $t^m/m!$ for arbitrary $t \in (0, 1)$. It is sufficient to show $\sum_{m=1}^n \rho_m ||\hat{b}_{m,n} - b_{m,N}||^2 = o_p(1)$ and $\sum_{m=1}^n \rho_m ||b_{m,N} - b_m||^2 = o_p(1)$, as $||\hat{b}_{m,n} - b_m||^2 \leqslant$

[8] Y_N is a random function $Y_N(\omega, u)$, where ω is in a probability space and u is a value in the interval $(0, 1)$. This is the setting when this lemma is applied in the proof of Lemma **9**. It is similar for $X_{i,N}$'s.

$2||\hat{b}_{m,n} - b_{m,N}||^2 + 2||b_{m,N} - b_m||^2$. Consider $\sum_{m=1}^n \rho_m ||b_{m,N} - b_m||^2 = o_p(1)$ first.

$$\sum_{m=1}^n \rho_m ||b_{m,N} - b_m||^2 = \sum_{m=1}^n \rho_m || \sum_{k=1}^\infty [\frac{1}{N} \nabla_{k,m} \ln L_N(\theta^0) - \nabla_{k,m} L_\infty(\theta^0)] 2^{-k} \sqrt{2} \cos(k\pi u)||^2$$
$$= \sum_{m=1}^n \rho_m \sum_{k=1}^\infty 2^{-2k} [\frac{1}{N} \nabla_{k,m} \ln L_N(\theta^0) - \nabla_{k,m} L_\infty(\theta^0)]^2.$$

By Lemmas 3, D.2 (1) and Assumption 18, there exists a constant $C_1 > 0$ that does not depend on m or N such that $[\mathrm{E}\, \frac{1}{N} \nabla_{k,m} \ln L_N(\theta^0) - \nabla_{k,m} L_\infty(\theta^0)]^2 \leqslant C_1 k^2$. Thus, given any $\epsilon > 0$, there exists a natural number K such that $\sum_{k=K}^\infty 2^{-2k} [\mathrm{E}\, \frac{1}{N} \nabla_{k,m} \ln L_N(\theta^0) - \nabla_{k,m} L_\infty(\theta^0)]^2 < \frac{\epsilon}{2}$. Hence,

$$\sum_{m=1}^n \rho_m \sum_{k=1}^\infty 2^{-2k} [\mathrm{E}\, \frac{1}{N} \nabla_{k,m} \ln L_N(\theta^0) - \nabla_{k,m} L_\infty(\theta^0)]^2$$
$$\leqslant \sum_{m=1}^n \rho_m \sum_{k=1}^{K-1} 2^{-2k} [\mathrm{E}\, \frac{1}{N} \nabla_{k,m} \ln L_N(\theta^0) - \nabla_{k,m} L_\infty(\theta^0)]^2 + \frac{\epsilon}{2} \sum_{m=1}^\infty \rho_m.$$

Similarly, because $\nabla_{k,m} = \nabla_{m,k}$, $[\frac{1}{N}\, \mathrm{E}\, \nabla_{k,m} \ln L_N(\theta^0) - \nabla_{k,m} L_\infty(\theta^0)]^2 \leqslant C_1 m^2$, there exists $m_0 \in \mathbb{N}$ such that $\sum_{m=m_0}^\infty \rho_m [\frac{1}{N}\, \mathrm{E}\, \nabla_{k,m} \ln L_N(\theta^0) - \nabla_{k,m} L_\infty(\theta^0)]^2 < \frac{\epsilon}{2}$. Because

$$\limsup_{N\to\infty} \sum_{m=1}^n \rho_m \sum_{k=1}^\infty 2^{-2k} [\frac{1}{N}\, \mathrm{E}\, \nabla_{k,m} \ln L_N(\theta^0) - \nabla_{k,m} L_\infty(\theta^0)]^2$$
$$\leqslant \limsup_{N\to\infty} \sum_{m=1}^{m_0-1} \rho_m \sum_{k=1}^{K-1} 2^{-2k} [\frac{1}{N}\, \mathrm{E}\, \nabla_{k,m} \ln L_N(\theta^0) - \nabla_{k,m} L_\infty(\theta^0)]^2 + \frac{\epsilon}{2} \sum_{m=1}^\infty \rho_m + \frac{\epsilon}{2} \leqslant \epsilon,$$

by Assumption 18, $\sum_{m=1}^n \rho_m ||b_{m,N} - b_m||^2 = o(1)$.

Next, we check the condition $\sum_{m=1}^n \rho_m ||\hat{b}_{m,n} - b_{m,N}||^2 = o_p(1)$. Whenever $\gamma \in [0,1]$, $||\gamma\hat{\theta}_n + (1-\gamma)\theta_n^0 - \theta^0||_3 = ||\gamma(\hat{\theta}_n - \theta^0) + (1-\gamma)(\theta_n^0 - \theta^0)||_3 \leqslant ||\hat{\theta}_n - \theta^0||_3 + ||\theta_n^0 - \theta^0||_3 = o_p(1)$ uniformly for γ by Theorem 1 on the consistency of $\hat{\theta}_n$ with $||\cdot||_3$ norm. Denote $\bar{\theta}_{kn} = \theta_n^0 + \gamma_k(\hat{\theta}_n - \theta_n^0)$. By Lemma D.2 (1), $|\frac{1}{N} \sum_{i=1}^N \nabla_{k,m}\, \mathrm{E}\, L_{i,N}(\theta^0)| \leqslant C_2 k$ for some constant C_2 that depends on neither k,

m nor N.

$$\sum_{m=1}^{n}\rho_m||\hat{b}_{m,n}-b_{m,N}||^2$$

$$=\sum_{m=1}^{n}\rho_m\sum_{k=1}^{n}2^{-2k}\left\{\frac{1}{N}\sum_{i=1}^{N}\left[\nabla_{k,m}L_{i,N}(\bar{\theta}_{kn})-\nabla_{k,m}\operatorname{E}L_{i,N}(\theta^0)\right]\right\}^2+$$

$$\sum_{m=1}^{n}\rho_m\sum_{k=n+1}^{\infty}2^{-2k}\left[\frac{1}{N}\sum_{i=1}^{N}\nabla_{k,m}\operatorname{E}L_{i,N}(\theta^0)\right]^2$$

$$\leqslant 2\sum_{m=1}^{n}\rho_m\sum_{k=1}^{\infty}2^{-2k}\left\{\frac{1}{N}\sum_{i=1}^{N}[\nabla_{k,m}L_{i,N}(\bar{\theta}_{kn})-\nabla_{k,m}L_{i,N}(\theta^0)]\right\}^2+$$

$$2\sum_{m=1}^{n}\rho_m\sum_{k=1}^{\infty}2^{-2k}\left\{\frac{1}{N}\sum_{i=1}^{N}[\nabla_{k,m}L_{i,N}(\theta^0)-\nabla_{k,m}\operatorname{E}L_{i,N}(\theta^0)]\right\}^2+\sum_{m=1}^{n}\rho_m\sum_{k=n+1}^{\infty}2^{-2k}C_2^2k^2$$

$$=2\sum_{m=1}^{n}\rho_m\sum_{k=1}^{\infty}2^{-2k}\left\{\frac{1}{N}\sum_{i=1}^{N}[\nabla_{k,m}L_{i,N}(\bar{\theta}_{kn})-\nabla_{k,m}L_{i,N}(\theta^0)]\right\}^2 1(||\hat{\theta}_n-\theta^0||_3+||\theta_n^0-\theta^0||_3\leqslant 1)+$$

$$2\sum_{m=1}^{n}\rho_m\sum_{k=1}^{\infty}2^{-2k}\left\{\frac{1}{N}\sum_{i=1}^{N}[\nabla_{k,m}L_{i,N}(\bar{\theta}_{kn})-\nabla_{k,m}L_{i,N}(\theta^0)]\right\}^2 1(||\hat{\theta}_n-\theta^0||_3+||\theta_n^0-\theta^0||_3>1)+$$

$$2\sum_{m=1}^{n}\rho_m\sum_{k=1}^{\infty}2^{-2k}\left\{\frac{1}{N}\sum_{i=1}^{N}[\nabla_{k,m}L_{i,N}(\theta^0)-\nabla_{k,m}\operatorname{E}L_{i,N}(\theta^0)]\right\}^2+o(1)$$

$$=2\sum_{m=1}^{n}\rho_m\sum_{k=1}^{\infty}2^{-2k}\left\{\frac{1}{N}\sum_{i=1}^{N}[\nabla_{k,m}L_{i,N}(\bar{\theta}_{kn})-\nabla_{k,m}L_{i,N}(\theta^0)]\right\}^2 1(||\hat{\theta}_n-\theta^0||_3+||\theta_n^0-\theta^0||_3\leqslant 1)$$

$$+2\sum_{m=1}^{n}\rho_m\sum_{k=1}^{\infty}2^{-2k}\left\{\frac{1}{N}\sum_{i=1}^{N}[\nabla_{k,m}L_{i,N}(\theta^0)-\nabla_{k,m}\operatorname{E}L_{i,N}(\theta^0)]\right\}^2+o_p(1).$$

By Lemmas 3 and D.2 (2) and Hölder's inequality[9],

$$\begin{aligned}&\left\|\left[\nabla_{k,m}L_{i,N}(\bar{\theta}_{kn})-\nabla_{k,m}L_{i,N}(\theta^0)\right]1(||\hat{\theta}_n-\theta^0||_3+||\theta_n^0-\theta^0||_3\leqslant 1)\right\|_{L^2}\\ \leqslant&C_3(k^2+m)\operatorname{E}^{1/4}\left[1(||\hat{\theta}_n-\theta^0||_3+||\theta_n^0-\theta^0||_3\leqslant 1)||\bar{\theta}_{kn}-\theta^0||_3^4\right]\\ \leqslant&C_3(k^2+m)\operatorname{E}^{1/4}\left\{1(||\hat{\theta}_n-\theta^0||_3+||\theta_n^0-\theta^0||_3\leqslant 1)\left[||\hat{\theta}_n-\theta^0||_3+||\theta_{n0}-\theta^0||_3\right]^4\right\}\\ =&C_3(k^2+m)\cdot o(1),\end{aligned}$$

[9] $||XY||_{L^2}\leqslant||X||_{L^4}||Y||_{L^4}$.

where the constant C_3 does not depend on k, m, i or N and $o(1)$ is uniformly in k and m. Accordingly, $\sum_{m=1}^{n}\rho_m\sum_{k=1}^{\infty}2^{-2k}\{\frac{1}{N}\sum_{i=1}^{N}[\nabla_{k,m}L_{i,N}(\bar{\theta}_{kn})-\nabla_{k,m}L_{i,N}(\theta_n^0)]\}^2 1(||\hat{\theta}_n-\theta^0||_3+||\theta_n^0-\theta^0||_3\leqslant 1)=o_p(1)$.

It remains to show that $\sum_{m=1}^{n}\rho_m\sum_{k=1}^{\infty}2^{-2k}\{\frac{1}{N}\sum_{i=1}^{N}[\nabla_{k,m}L_{i,N}(\theta^0)-\mathrm{E}\,\nabla_{k,m}L_{i,N}(\theta^0)]\}^2=o_p(1)$ and it suffices to prove that its expectation is $o(1)$. By Proposition 3 and Lemma E.1, there exists a constant C_4 such that $\mathrm{var}(\frac{1}{N}\sum_{i=1}^{N}\nabla_{j,k}L_{i,N}(\theta_0))\leqslant\frac{1}{N}C_4(j^2+k^2)$ for all $j\in\mathbb{N}$ and $k\in\mathbb{N}$. Hence, as $N\to\infty$,

$$\mathrm{E}\left[\sum_{m=1}^{n}\rho_m\sum_{k=1}^{\infty}2^{-2k}\{\frac{1}{N}\sum_{i=1}^{N}[\nabla_{k,m}L_{i,N}(\theta^0)-\mathrm{E}\,L_{i,N}(\theta^0)]\}^2\right]\leqslant\frac{C_4}{N}\sum_{m=1}^{\infty}\rho_m\sum_{k=1}^{\infty}2^{-2k}(j^3+k^3)\to 0.$$

Therefore, $\mathrm{E}\sum_{m=1}^{n}\rho_m\sum_{k=1}^{\infty}2^{-2k}\{\frac{1}{N}\sum_{i=1}^{N}[\nabla_{k,m}L_{i,N}(\theta^0)-\mathrm{E}\,L_{i,N}(\theta^0)]\}^2=o(1)$.

Finally, we need to check condition (c) in Lemma E.2: $\liminf_{N\to\infty}||\sum_{m=2+K^0}^{n}\rho_m b_m||>0$. If for all $t\in(0,1)$, we have $\liminf_{N\to\infty}||\sum_{m=2+K^0}^{n}\frac{t^m}{m!}b_m||=||\sum_{m=2+K^0}^{\infty}\frac{t^m}{m!}b_m||=0$, then $\sum_{k=1}^{\infty}2^{-2k}[\sum_{m=2+K^0}^{\infty}\frac{t^m}{m!}\nabla_{k,m}L_\infty(\theta^0)]^2=0$. Thus, $\sum_{m=2+K^0}^{\infty}\frac{t^m}{m!}\nabla_{k,m}L_\infty(\theta^0)\equiv 0$ for all $t\in(0,1)$ and for all $k\in\mathbb{N}$. Differentiating this equation with respect to t for $m\geqslant 2+K^0$ times, we obtain $\nabla_{k,m}L_\infty(\theta^0)=0$ for all $k\in\mathbb{N}$, contradicting Assumption 19. So there exists a $t\in(0,1)$ and a corresponding ρ_m such that condition (c) holds. □

Proof of Proposition 4: (1) With the information identity $\mathrm{E}[\nabla_k\ln L_N(\theta^0)\cdot\nabla_m\ln L_N(\theta^0)=-\mathrm{E}[\nabla_{k,m}\ln L_N(\theta^0)]$, it is equivalent to show that

$$\sup_{u_1,u_2\in[0,1]}\left|\mathrm{E}\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\left[\nabla_{k,m}L_\infty(\theta^0)-\frac{1}{N}\nabla_{k,m}\ln L_N(\theta^0)\right]\eta_k(u_1)\eta_m(u_2)\right|=o(1), \tag{E.4}$$

$$\sup_{u_1,u_2\in[0,1]}\left|\mathrm{E}(\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}-\sum_{k=1}^{n}\sum_{m=1}^{n})\left[\frac{1}{N}\nabla_{k,m}\ln L_N(\theta^0)\eta_k(u_1)\eta_m(u_2)\right]\right|=o(1), \tag{E.5}$$

$$\sup_{u_1,u_2\in[0,1]}\left|\sum_{k=1}^{n}\sum_{m=1}^{n}\frac{1}{N}\left[\nabla_{k,m}\ln L_N(\theta^0)-\mathrm{E}\,\nabla_{k,m}\ln L_N(\theta^0)\right]\eta_k(u_1)\eta_m(u_2)\right|=o_p(1), \tag{E.6}$$

and

$$\sup_{u_1,u_2\in[0,1]}\left|\frac{1}{N}\sum_{k=1}^{n}\sum_{m=1}^{n}\left[\nabla_{k,m}\ln L_N(\hat{\theta}_n)-\nabla_{k,m}\ln L_N(\theta^0)\right]\eta_k(u_1)\eta_m(u_2)\right|=o_p(1). \tag{E.7}$$

For Eq. (E.4), by Lemma D.2 (1), there exists some constant $C_1>0$ such that $|\frac{1}{N}\nabla_{k,m}\ln L_N(\theta^0)|\leqslant$

$C_1 m A_N$ for all k and m, where

$$A_N \equiv \frac{1}{N} \max\{N, \sum_i (1+|z_{i,N}|) \max(|w_{i\cdot,N} Y_N|, \max_{1\leqslant j\leqslant K^0} |x_{ij,N}|),$$

$$\sum_i (1+z_{i,N}^2) \max[(w_{i\cdot,N}Y_N)^2, \max_{1\leqslant k\leqslant K^0} |x_{ik,N}\cdot w_{i\cdot,N}Y_N|, \max_{1\leqslant j,k\leqslant K^0} |x_{ik,N}\cdot x_{ij,N}|]\}.$$

By Lemma 3, A_N is uniformly L_p bounded in N, and $A_N = O_p(1)$. Thus, under Assumption 18, $|\nabla_{k,m} L_\infty(\theta^0)| \leqslant C_1 C_2 m$ for all k and m, where $C_2 \equiv \sup_N \mathrm{E}\, A_N < \infty$. Hence, for any $\epsilon > 0$, there exists $K_\epsilon \in \mathbb{N}$ that does not depend on N such that

$$\begin{aligned}
&\sup_{u_1,u_2\in[0,1]} \left| \left(\sum_{k=1}^{\infty}\sum_{m=1}^{\infty} - \sum_{k=1}^{K_\epsilon}\sum_{m=1}^{K_\epsilon} \right) \left[\nabla_{k,m} L_\infty(\theta^0) - \frac{1}{N}\nabla_{k,m} \mathrm{E} \ln L_N(\theta^0) \right] \eta_k(u_1)\eta_m(u_2) \right| \\
\leqslant &\sup_{u_1,u_2\in[0,1]} \left(\sum_{k=K_\epsilon+1}^{\infty}\sum_{m=1}^{\infty} + \sum_{k=1}^{\infty}\sum_{m=K_\epsilon+1}^{\infty} \right) \left| \left[\nabla_{k,m} L_\infty(\theta^0) - \frac{1}{N}\nabla_{k,m} \mathrm{E} \ln L_N(\theta^0) \right] \eta_k(u_1)\eta_m(u_2) \right| \\
\leqslant &\left(\sum_{k=K_\epsilon+1}^{\infty}\sum_{m=1}^{\infty} + \sum_{k=1}^{\infty}\sum_{m=K_\epsilon+1}^{\infty} \right) \sum_{k=K_\epsilon+1}^{\infty}\sum_{m=1}^{\infty} 2C_1C_2 m\cdot 2^{1-k-m} < \epsilon.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\limsup_{N\to\infty} \sup_{u_1,u_2\in[0,1]} \left| \sum_{k=1}^{K_\epsilon}\sum_{m=1}^{K_\epsilon} \left[\nabla_{k,m} L_\infty(\theta^0) - \frac{1}{N}\nabla_{k,m} \mathrm{E} \ln L_N(\theta^0) \right] \eta_k(u_1)\eta_m(u_2) \right| \\
\leqslant &\limsup_{N\to\infty} \sum_{k=1}^{K_\epsilon}\sum_{m=1}^{K_\epsilon} \left| \nabla_{k,m} L_\infty(\theta^0) - \frac{1}{N}\nabla_{k,m} \mathrm{E} \ln L_N(\theta^0) \right| 2^{1-k-m} = 0.
\end{aligned}$$

Hence, Eq. (E.4) holds.

For Eq. (E.5) follows because

$$\begin{aligned}
&\sup_{u_1,u_2\in[0,1]} |\,\mathrm{E}(\sum_{k=1}^{\infty}\sum_{m=1}^{\infty} - \sum_{k=1}^{n}\sum_{m=1}^{n})[\frac{1}{N}\nabla_{k,m}\ln L_N(\theta^0)\eta_k(u_1)\eta_m(u_2)]| \\
\leqslant &\sum_{k=n+1}^{\infty}\sum_{m=1}^{\infty} \mathrm{E}\,|\frac{1}{N}\nabla_{k,m}\ln L_N(\theta^0)|2^{1-k-m} + \sum_{k=1}^{n}\sum_{m=n+1}^{\infty} \mathrm{E}\,|\frac{1}{N}\nabla_{k,m}\ln L_N(\theta^0)|2^{1-k-m} \\
\leqslant &C_1C_2 \left(\sum_{k=n+1}^{\infty}\sum_{m=1}^{\infty} m2^{1-k-m} + \sum_{k=1}^{\infty}\sum_{m=n+1}^{\infty} m2^{1-k-m} \right) = C_1C_2\left(\frac{3}{2^n} + \frac{n+2}{2^{n-1}}\right) = o_p(1).
\end{aligned}$$

For Eq. (E.6), because $A_N = O_p(1)$ and

$$\begin{aligned}
&\sup_{u_1,u_2\in[0,1]}\left|\left(\sum_{k=1}^{n}\sum_{m=1}^{n}-\sum_{k=1}^{K'_\epsilon}\sum_{m=1}^{K'_\epsilon}\right)\frac{1}{N}[\nabla_{k,m}\ln L_N(\theta^0)-\mathrm{E}\,\nabla_{k,m}\ln L_N(\theta^0)]\eta_k(u_1)\eta_m(u_2)\right|\\
\leqslant&\sup_{u_1,u_2\in[0,1]}(\sum_{k=K'_\epsilon+1}^{\infty}\sum_{m=1}^{\infty}+\sum_{k=1}^{\infty}\sum_{m=K'_\epsilon+1}^{\infty})\left|\frac{1}{N}\left[\nabla_{k,m}\ln L_N(\theta^0)-\mathrm{E}\,\nabla_{k,m}\ln L_N(\theta^0)\right]\eta_k(u_1)\eta_m(u_2)\right|\\
\leqslant&\sum_{k=K'_\epsilon+1}^{\infty}\sum_{m=1}^{\infty}2^{1-k-m}C_1m(A_N+C_2)=2C_1(A_N+C_2)\sum_{k=K'_\epsilon+1}^{\infty}2^{-k}\sum_{m=1}^{\infty}2^{-m}m\\
=&4C_1(A_N+C_2)2^{-K'_\epsilon}=4C_1(A_N+C_2)2^{-K'_\epsilon}(K'_\epsilon+4),
\end{aligned}$$

there exists a large enough constant $K'_\epsilon \in \mathbb{N}$ that does not depend on N such that

$$P(\sup_{u_1,u_2\in[0,1]}|(\sum_{k=1}^{n}\sum_{m=1}^{n}-\sum_{k=1}^{K'_\epsilon}\sum_{m=1}^{K'_\epsilon})\frac{1}{N}[\nabla_{k,m}\ln L_N(\theta^0)-\mathrm{E}\,\nabla_{k,m}\ln L_N(\theta^0)]\eta_k(u_1)\eta_m(u_2)]|>\frac{\epsilon}{2})<\frac{\epsilon}{2}. \tag{E.8}$$

For each pair (k,m), $\{\nabla_{k,m}\ln L_{i,N}(\theta^0)\}_{i=1}^N$ satisfies the WLLN in JP (2012) as it is UG L_2-NED from Proposition 3, and uniformly L_p bounded for any $p>1$ from Lemma D.2 (1). Hence,

$$\begin{aligned}
&P(\sup_{u_1,u_2\in[0,1]}|\sum_{k=1}^{n}\sum_{m=1}^{n}\frac{1}{N}[\nabla_{k,m}\ln L_N(\theta^0)-\mathrm{E}\,\nabla_{k,m}\ln L_N(\theta^0)]\eta_k(u_1)\eta_m(u_2)]|>\epsilon)\\
\leqslant&P\left(\sup_{u_1,u_2\in[0,1]}|\sum_{k=1}^{K'_\epsilon}\sum_{m=1}^{K'_\epsilon}\frac{1}{N}[\nabla_{k,m}\ln L_N(\theta^0)-\mathrm{E}\,\nabla_{k,m}\ln L_N(\theta^0)]\eta_k(u_1)\eta_m(u_2)]|>\frac{\epsilon}{2}\right)+\\
&P\left(\sup_{u_1,u_2\in[0,1]}|(\sum_{k=1}^{n}\sum_{m=1}^{n}-\sum_{k=1}^{K'_\epsilon}\sum_{m=1}^{K'_\epsilon})\frac{1}{N}[\nabla_{k,m}\ln L_N(\theta^0)-\mathrm{E}\,\nabla_{k,m}\ln L_N(\theta^0)]\eta_k(u_1)\eta_m(u_2)]|>\frac{\epsilon}{2}\right)\\
\leqslant&\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon,
\end{aligned}$$

for large enough N.

For Eq. (E.7), from Lemma D.2 (2), there exists a sequence of random variables $\bar{A}_N$, which is uniformly L_p bounded for any $p \geqslant 1$, such that $|\frac{1}{N}\nabla_{k,m}\ln L_N(\theta^1)-\frac{1}{N}\nabla_{k,m}\ln L_N(\theta^2)| \leqslant \bar{A}_N(k+m+k^2)||\theta^1-\theta^2||_2$ for any natural numbers k and m, when θ^1 and θ^2 are close to each other.

Therefore,

$$\begin{aligned}
&\sup_{u_1,u_2\in[0,1]}|\frac{1}{N}\sum_{k=1}^{n}\sum_{m=1}^{n}[\nabla_{k,m}\ln L_N(\hat{\theta}_n)-\nabla_{k,m}\ln L_N(\theta^0)]\eta_k(u_1)\eta_m(u_2)|\\
\leqslant&\sum_{k=1}^{n}\sum_{m=1}^{n}|\frac{1}{N}[\nabla_{k,m}\ln L_N(\hat{\theta}_n)-\nabla_{k,m}\ln L_N(\theta^0)]|2^{1-k-m}+o_p(1)\\
\leqslant&\bar{A}_N\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}(k+m+k^2)2^{1-k-m}\cdot||\hat{\theta}_n-\theta^0||_2+o_p(1)=o_p(1).
\end{aligned}$$

(2) By Lemma 9 and Assumption 20, $0<\int_0^1 a(u)a(u)'du<\infty$ and $\mathrm{plim}_{N\to\infty}\int_0^1 \bar{a}_n(u)\bar{a}_n(u)'du=\int_0^1 a(u)a(u)'du$. And from Lemma 8, $\sup_{0\leqslant u_1,u_2\leqslant 1}|\Gamma(u_1,u_2)|<\infty$. Then, the conclusion holds because

$$\begin{aligned}
&\left|\int_0^1\int_0^1 \bar{a}_{i,n}(u_1)\hat{\Gamma}_n(u_1,\,u_2)\bar{a}_{j,n}(u_2)du_1du_2-\int_0^1\int_0^1 a_i(u_1)\Gamma(u_1,\,u_2)a_j(u_2)du_1du_2\right|\\
\leqslant&\left|\int_0^1\int_0^1 \bar{a}_{i,n}(u_1)\hat{\Gamma}_n(u_1,\,u_2)\bar{a}_{j,n}(u_2)du_1du_2-\int_0^1\int_0^1 \bar{a}_{i,n}(u_1)\Gamma(u_1,\,u_2)\bar{a}_{j,n}(u_2)du_1du_2\right|+\\
&\left|\int_0^1\int_0^1 \bar{a}_{i,n}(u_1)\Gamma(u_1,\,u_2)\bar{a}_{j,n}(u_2)du_1du_2-\int_0^1\int_0^1 a_i(u_1)\Gamma(u_1,\,u_2)\bar{a}_{j,n}(u_2)du_1du_2\right|+\\
&\left|\int_0^1\int_0^1 a_i(u_1)\Gamma(u_1,\,u_2)\bar{a}_{j,n}(u_2)du_1du_2-\int_0^1\int_0^1 a_i(u_1)\Gamma(u_1,\,u_2)a_j(u_2)du_1du_2\right|\\
\leqslant&\sup_{u_1,u_2\in[0,1]}|\hat{\Gamma}_n(u_1,\,u_2)-\Gamma(u_1,\,u_2)|\cdot\int_0^1\int_0^1|\bar{a}_{i,n}(u_1)\bar{a}_{j,n}(u_2)|du_1du_2+\\
&\sup_{0\leqslant u_1,u_2\leqslant 1}|\Gamma(u_1,u_2)|\cdot\int_0^1\int_0^1|[\bar{a}_{i,n}(u_1)-a_i(u_1)]\bar{a}_{j,n}(u_2)|du_1du_2+\\
&\sup_{0\leqslant u_1,u_2\leqslant 1}|\Gamma(u_1,u_2)|\cdot\int_0^1\int_0^1|a_i(u_1)[\bar{a}_{j,n}(u_2)-a_j(u_2)]|du_1du_2\\
\leqslant&o_p(1)+\sup_{0\leqslant u_1,u_2\leqslant 1}|\Gamma(u_1,u_2)|\cdot\sqrt{\int_0^1[\bar{a}_{i,n}(u_1)-a_i(u_1)]^2du_1}\sqrt{\int_0^1\bar{a}_{j,n}(u_2)^2du_2}\\
&+\sup_{0\leqslant u_1,u_2\leqslant 1}|\Gamma(u_1,u_2)|\cdot\sqrt{\int_0^1 a_i(u_1)^2du_1}\sqrt{\int_0^1[\bar{a}_{j,n}(u_2)-a_j(u_2)]^2du_2}=o_p(1),
\end{aligned}$$

where the third inequality is based on the Cauchy–Schwarz inequality. □

Proof of Proposition 5: For any $p(u)=\sum_{k=1}^n p_k\eta_k(u)=\sum_{k=1}^n p_k\omega_k\chi_k(u)=p\chi(u)$, where $p=(p_1,\cdots,p_n)\Lambda_n$ and $\chi(u)=(\chi_1(u),\cdots,\chi_n(u))'$, the vector of coefficients p can characterize $p(u)$. Thus, $\bar{b}_{m,n}=-(\hat{H}_{1m,n},\cdots,\hat{H}_{nm,n})\Lambda_n$ characterizes the function $\bar{b}_{m,n}(u)$. $\bar{b}_{m,n}$ can be decomposed into $(\bar{b}'_{1,n}, \cdots, \bar{b}'_{1+K^0,n}) = -\Lambda_n \hat{H}_{n,(1:K^0+1)}$ and $(\bar{b}'_{K^0+1,n}, ..., \bar{b}'_{n,n}) = -\Lambda_n \hat{H}_{n,(K^0+2:n)}$, where $\hat{H}_{n,(1:K^0+1)}$ and $\hat{H}_{n,(K^0+2:n)}$ are the first K^0+1 columns and the rest columns of $\hat{H}_n$. Since $\bar{a}_{m,n}$ is the residual of projection of $\bar{b}_{m,n}$ on $\bar{b}_{K^0+1,n}$, ..., $\bar{b}_{n,n}$ when $1 \leqslant m \leqslant K^0+1$,

$$\begin{aligned}
&\bar{a}'_n \equiv (\bar{a}'_{1,n}, \cdots, \bar{a}'_{1+K^0,n}) \\
=& -[I_n - \Lambda_n \hat{H}_{n,(K^0+2:n)}(\hat{H}'_{n,(K^0+2:n)}\Lambda_n^2 \hat{H}_{n,(K^0+2:n)})^{-1}\hat{H}'_{n,(K^0+2:n)}\Lambda_n]\Lambda_n \hat{H}_{n,(1:K^0+1)},
\end{aligned}$$

i.e., $\bar{a}_n(u) = -\hat{H}'_{n,(1:K^0+1)}[I_n - \Lambda_n^2 \hat{H}_{n,(K^0+2:n)}(\hat{H}'_{n,(K^0+2:n)}\Lambda_n^2 \hat{H}_{n,(K^0+2:n)})^{-1}\hat{H}'_{n,(K^0+2:n)}]\eta(u)$. Denote

$$\begin{aligned}
&A_{K^0+1,K^0+1,n} \equiv \int_0^1 \bar{a}_n(u)\bar{a}_n(u)'du \equiv \bar{a}_n \bar{a}'_n \\
=&\hat{H}'_{n,(1:K^0+1)}\Lambda_n[I_n - \Lambda_n \hat{H}_{n,(K^0+2:n)}(\hat{H}'_{n,(K^0+2:n)}\Lambda_n^2 \hat{H}_{n,(K^0+2:n)})^{-1}\hat{H}'_{n,(K^0+2:n)}\Lambda_n]\Lambda_n \hat{H}_{n,(1:K^0+1)}.
\end{aligned}$$

Because $\hat{\Gamma}_n(u_1, u_2) = -\eta(u_1)'\hat{H}_n \eta(u_2)$,

$$\begin{aligned}
&\int_0^1 \int_0^1 \bar{a}_n(u_1)\hat{\Gamma}_n(u_1, u_2)\bar{a}_n(u_2)'du_1 du_2 \\
=& -\hat{H}'_{n,(1:K^0+1)}[I_n - \Lambda_n^2 \hat{H}_{n,(K^0+2:n)}(\hat{H}'_{n,(K^0+2:n)}\Lambda_n^2 \hat{H}_{n,(K^0+2:n)})^{-1}\hat{H}'_{n,(K^0+2:n)}]\Lambda_n^2 \hat{H}_n \\
&\Lambda_n^2[I_n - \hat{H}_{n,(K^0+2:n)}(\hat{H}'_{n,(K^0+2:n)}\Lambda_n^2 \hat{H}_{n,(K^0+2:n)})^{-1}\hat{H}'_{n,(K^0+2:n)}\Lambda_n^2]\hat{H}_{n,(1:K^0+1)} \\
=& -\hat{H}'_{n,(1:K^0+1)}[I_n - \Lambda_n^2 \hat{H}_{n,(K^0+2:n)}(\hat{H}'_{n,(K^0+2:n)}\Lambda_n^2 \hat{H}_{n,(K^0+2:n)})^{-1}\hat{H}'_{n,(K^0+2:n)}] \\
&\Lambda_n^2(\hat{H}_{n,(1:K^0+1)}, \hat{H}_{n,(K^0+2:n)}) \cdot \hat{H}_n^{-1} \cdot (\hat{H}_{n,(1:K^0+1)}, \hat{H}_{n,(K^0+2:n)})' \\
&\Lambda_n^2[I_n - \hat{H}_{n,(K^0+2:n)}(\hat{H}'_{n,(K^0+2:n)}\Lambda_n^2 \hat{H}_{n,(K^0+2:n)})^{-1}\hat{H}'_{n,(K^0+2:n)}\Lambda_n^2]\hat{H}_{n,(1:K^0+1)} \\
=& -A_{K^0+1,K^0+1,n}(I_{K^0+1}, O_{K^0+1,n-K^0-1})\hat{H}_n^{-1}(I_{K^0+1}, O_{K^0+1,n-K^0-1})A'_{K^0+1,K^0+1,n}.
\end{aligned}$$

Thus, the conclusion holds. □

References

[1] Abramowitz, M., Stegun. I, 1967. Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables. US Government Printing Office.

[2] Allouch, N., 2012. On the Private Provision of Public Goods on Networks. Working paper, University of London.

[3] Amaral, P., Anselin, L., 2013. Finite sample properties of Moran's I test for spatial autocorrelation in Tobit models. Papers in Regional Science, 93(4), 773-781.

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

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

[6] Andrews, D. W., 1992. Generic uniform convergence. Econometric theory, 8(02), 241-257.

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

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

[9] Bierens, H. J., 2014. Consistency and asymptotic normality of sieve ML estimators under low-level conditions. Econometric Theory 30, 1021-1076.

[10] Billingsley, P., 1999. Convergence of Probability Measures. Second edition. John Wiley & Sons. New York.

[11] Bramoullé, Y., Kranton, R., D'Amours, M., 2014. Strategic interaction and networks. The American Economic Review 104, 898-930.

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

[13] Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models, Handbook of Econometrics, Vol. 6. North-Holland, New York, 5549-5632.

[14] Dedecker, J., 2001. Exponential inequalities and functional central limit theorems for random fields. Probability and Statistics 5, 77-104.

[15] Delyon, B., 2009. Exponential inequalities for sums of weakly dependent variables. Electronic Journal of Probability 14, 752-779.

[16] 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. Working paper, University of Rennes I.

[17] Doukhan, P., 1994. Mixing: Properties and Examples. Springer, New York.

[18] Doukhan, P. and Louhichi, S., 1999. A new weak dependence condition and applications to moment inequalities. Stochastic Processes and their Applications 84, 313-342.

[19] Folland, G. B., 1999. Real Analysis: Modern Techniques and Their Applications (Vol. 361). John Wiley & Sons, New York.

[20] Gallant, A. R., Nychka, D. W., 1987. Semi-nonparametric maximum likelihood estimation. Econometrica 55, 363-390.

[21] Jackson, M., Zenou, Y., 2014. Games on networks, in: Young, P., Zamir, S., (Eds.), Handbook of Game Theory, Vol. 4. North-Holland, Amsterdam, pp. 95-163.

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

[23] Jenish, N., Prucha, I. R., 2009. Central limit theorems and uniform laws of large numbers for arrays of random fields. Journal of Econometrics 150, 86-98.

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

[25] Ko, M.H., 2013. The complete convergence for weighted sums of dependent random fields. Journal of Inequalities and Applications 2013, 1-8.

[26] Kreyszig, E., 1978. Introductory Functional Analysis with Applications. John Wiley & Sons, New York.

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

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

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

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

[31] Panchenko, Dmitry, 2007. 18.465 Topics in Statistics: Statistical Learning Theory, Spring 2007. (MIT OpenCourseWare: Massachusetts Institute of Technology), http://ocw.mit.edu/courses/mathematics/18-465-topics-in-statistics-statistical-learning-theory-spring-2007.

[32] Olver, F. W., 2010. NIST Handbook of Mathematical Functions. Cambridge University Press, New York.

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

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

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

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

[37] Royden, H., Fitzpatrick, P., Real Analysis. 2010. China Machine Press, Beijing.

[38] Saulis, L., Statulevicius, V. A., 1991. Limit Theorems for Large Deviations. Kluwer Academic Publishers, Norwell, MA.

[39] Steele, J. M., 2004. The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities. Cambridge University Press.

[40] U.S. Dept. of Commerce, Bureau of the Census, 1992. Contiguous county file, 1991: [United States]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research.

[41] White, H., Wooldridge, J., 1991. Some results on sieve estimation with dependent observations. in: W. Barnett, J. Powell and G. Tauchen, Non-parametric and Semi-parametric Methods in Econometrics and Statistics. Cambridge University Press, Cambridge, 459-493.

[42] Xu, X., Lee, L., 2015a. A spatial autoregressive model with a nonlinear transformation of the dependent variable. Journal of Econometrics 186, 1-18.

[43] Xu, X., Lee, L., 2015b. Maximum likelihood estimation of a spatial autoregressive Tobit model. Journal of Econometrics 188, 264-280.

Table 1: Estimation Results under Mixed Normally Distributed Disturbances											
sample size			No Symmetry					standard Logit		Symmetry	
			PMLE	AIC	BIC	5	10	AIC	BIC	AIC	BIC
200	λ	mean	0.4187	0.5062	0.5350	0.5799	0.4953	0.4849	0.4785	0.4926	0.4913
		std	0.1480	0.0654	0.0664	0.0714	0.0663	0.0633	0.0602	0.0624	0.0639
		RMSE	0.1688	0.0657	0.0751	0.1071	0.0665	0.0651	0.0639	0.0628	0.0644
		med	0.4326	0.5050	0.5363	0.5828	0.4977	0.4855	0.4788	0.4923	0.4903
	β_1	mean	-0.6240							-0.9499	-0.9189
		std	0.2277							0.1309	0.1334
		RMSE	0.4395							0.1401	0.1561
		med	-0.6108							-0.9515	-0.9184
	β_2	mean	1.5930	2.0039	2.0324	2.1093	1.9960	1.9755	1.9418	1.9768	1.9592
		std	0.3097	0.1161	0.1215	0.1405	0.1144	0.1080	0.0978	0.1084	0.1106
		RMSE	0.5144	0.1161	0.1258	0.1780	0.1145	0.1107	0.1138	0.1108	0.1179
		med	1.5915	1.9999	2.0368	2.0993	1.9932	1.9695	1.9451	1.9792	1.9624
		AIC	562.5	420.6				421.4		420.9	
		BIC	575.7		450.9				445.8		445.8
500	λ	mean	0.4220	0.5064	0.5279	0.6083	0.5011	0.4897	0.4755	0.4932	0.4926
		std	0.1147	0.0520	0.0531	0.0621	0.0522	0.0495	0.0463	0.0475	0.0477
		RMSE	0.1387	0.0524	0.0600	0.1248	0.0522	0.0505	0.0524	0.0480	0.0483
		med	0.4287	0.5060	0.5272	0.6057	0.5005	0.4885	0.4752	0.4939	0.4932
	β_1	mean	-0.5779							-0.9272	-0.9218
		std	0.1553							0.0904	0.0922
		RMSE	0.4497							0.1161	0.1209
		med	-0.5750							-0.9305	-0.9241
	β_2	mean	1.5507	2.0017	2.0140	2.0760	1.9995	1.9883	1.9577	1.9719	1.9692
		std	0.1915	0.0768	0.0795	0.0943	0.0756	0.0766	0.0718	0.0730	0.0738
		RMSE	0.4884	0.0768	0.0807	0.1211	0.0756	0.0775	0.0833	0.0782	0.0799
		med	1.5521	2.0012	2.0117	2.0741	2.0005	1.9888	1.9567	1.9718	1.9694
		AIC	1376.9	1029.6				1028.9		1029.4	
		BIC	1393.8		1070.0				1063.4		1062.9
1000	λ	mean	0.4458	0.5041	0.5131	0.5976	0.5013	0.4967	0.4910	0.5099	0.5100
		std	0.0643	0.0333	0.0338	0.0390	0.0335	0.0330	0.0342	0.0308	0.0308
		RMSE	0.0841	0.0335	0.0362	0.1051	0.0336	0.0332	0.0353	0.0324	0.0324
		med	0.4500	0.5041	0.5119	0.5949	0.5006	0.4965	0.4896	0.5102	0.5102
	β_1	mean	-0.5902							-0.9327	-0.9325
		std	0.0955							0.0638	0.0637
		RMSE	0.4208							0.0927	0.0928
		med	-0.5909							-0.9353	-0.9350
	β_2	mean	1.5868	1.9960	1.9987	2.0380	1.9963	1.9906	1.9661	1.9599	1.9598
		std	0.1377	0.0544	0.0561	0.0669	0.0541	0.0581	0.0568	0.0546	0.0546
		RMSE	0.4355	0.0546	0.0561	0.0770	0.0542	0.0589	0.0662	0.0677	0.0678
		med	1.5879	1.9960	1.9970	2.0386	1.9954	1.9932	1.9664	1.9587	1.9586
		AIC	2816.1	2131.8				2134.4		2134.6	
		BIC	2835.8		2181.9				2179.9		2173.9

$\lambda_0 = 0.5$, $\beta_{10} = -1$, $\beta_{20} = 2$. Mixed Normal distribution: half probability $N(8/\sqrt{17}, 4/17)$, half probability $N(-8/\sqrt{17}, 4/17)$
PMLE: parametric MLE, i.e., estimate the model under the normal distribution assumption of $\epsilon_{i,N}$

Table 2: Estimation Results under Laplace Distribution Disturbances

sample size			No Symmetry					standard Logit		symmetry	
			PMLE	AIC	BIC	5	10	AIC	BIC	AIC	BIC
200	λ	mean	0.5000	0.4907	0.4869	0.4842	0.4893	0.4859	0.4985	0.4806	0.4859
		std	0.1545	0.1502	0.1420	0.1542	0.1805	0.1527	0.1366	0.1554	0.1413
		RMSE	0.1545	0.1504	0.1426	0.1550	0.1808	0.1533	0.1367	0.1566	0.1420
		med	0.5167	0.5079	0.5029	0.4986	0.5064	0.5033	0.5185	0.4953	0.5037
	β_1	mean	-1.1964							-1.0022	-1.0089
		std	0.2850							0.2586	0.2381
		RMSE	0.3461							0.2587	0.2382
		med	-1.1813							-0.9785	-0.9922
	β_2	mean	2.2322	2.0387	2.0247	2.0199	2.0600	2.0351	2.0649	2.0294	2.0255
		std	0.3628	0.3466	0.3149	0.3482	0.4033	0.3591	0.2946	0.3545	0.3111
		RMSE	0.4308	0.3487	0.3159	0.3487	0.4078	0.3608	0.3016	0.3557	0.3122
		med	2.2238	2.0279	2.0039	2.0091	2.0246	2.0211	2.0852	2.0099	2.0157
		AIC	525.2	511.8				512.2		511.4	
		BIC	538.4		526.6				527.2		526.6
500	λ	mean	0.5125	0.4942	0.4948	0.4919	0.4990	0.4962	0.5116	0.4879	0.4912
		std	0.1154	0.1113	0.1042	0.1084	0.1275	0.1106	0.1002	0.1065	0.1033
		RMSE	0.1160	0.1114	0.1043	0.1087	0.1275	0.1106	0.1009	0.1072	0.1037
		med	0.5159	0.4990	0.5003	0.4979	0.4997	0.5009	0.5155	0.4952	0.4942
	β_1	mean	-1.2273							-0.9916	-1.0029
		std	0.1845							0.1558	0.1495
		RMSE	0.2928							0.1560	0.1496
		med	-1.2297							-0.9889	-0.9993
	β_2	mean	2.2472	2.0176	2.0144	2.0025	2.0275	2.0233	2.0571	2.0042	2.0096
		std	0.2415	0.2117	0.2024	0.2132	0.2380	0.2261	0.1930	0.2132	0.2004
		RMSE	0.3456	0.2125	0.2029	0.2132	0.2396	0.2273	0.2012	0.2132	0.2006
		med	2.2480	2.0060	2.0100	1.9964	2.0269	2.0279	2.0608	1.9970	2.0034
		AIC	1278.2	1246.0				1247.3		1245.4	
		BIC	1295.0		1265.1				1266.3		1264.8
1000	λ	mean	0.5139	0.5173	0.5282	0.5140	0.5090	0.5206	0.5361	0.5093	0.5189
		std	0.0775	0.0671	0.0634	0.0658	0.0723	0.0672	0.0607	0.0668	0.0641
		RMSE	0.0788	0.0693	0.0694	0.0673	0.0729	0.0703	0.0707	0.0675	0.0669
		med	0.5174	0.5189	0.5320	0.5170	0.5109	0.5243	0.5368	0.5117	0.5211
	β_1	mean	-1.2394							-1.0069	-1.0200
		std	0.1229							0.0991	0.0997
		RMSE	0.2691							0.0994	0.0997
		med	-1.2357							-1.0031	-1.0172
	β_2	mean	2.2380	2.0015	1.9997	1.9933	1.9996	2.0027	2.0368	1.9958	1.9954
		std	0.1811	0.1526	0.1471	0.1475	0.1678	0.1618	0.1456	0.1498	0.1460
		RMSE	0.2991	0.1526	0.1471	0.1477	0.1678	0.1618	0.1501	0.1499	0.1461
		med	2.2236	1.9968	1.9942	1.9903	1.9953	1.9984	2.0350	1.9910	1.9862
		AIC	2614.3	2551.7				2554.3		2550.8	
		BIC	2634.0		2574.8				2576.8		2574.2

$\lambda_0 = 0.5$, $\beta_{10} = -1$, $\beta_{20} = 2$. Laplace Distribution: expectation = 0, standard deviation = 2
PMLE: parametric MLE, i.e., estimate the model under the normal distribution assumption of $\epsilon_{i,N}$

Table 3: Frequency of Number of Basis Functions Used in the Experiment

	Mixed Normal						Laplace					
sample size	200		500		1000		200		500		1000	
#(sieves)	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
2	0	0	0	0	0	0	478	933	431	939	298	924
3	0	3	0	0	0	0	118	45	62	27	55	22
4	0	0	0	0	0	0	135	19	208	32	304	52
5	120	414	12	137	0	3	75	0	98	2	110	2
6	163	331	21	182	4	58	51	1	50	0	49	0
7	141	97	142	221	33	147	47	0	50	0	42	0
8	399	146	525	425	484	696	39	1	52	0	60	0
9	75	6	100	28	150	63	29	1	30	0	44	0
10	102	3	200	7	329	33	28	0	19	0	38	0

Table 4: The difference by using 10 and 15 sieves

	λ_0	mean(—$\hat{\lambda}_{10} - \hat{\lambda}_{15}$—)	$\sqrt{\text{mean}(\hat{\lambda}_{10} - \hat{\lambda}_{15})^2}$	β_{20}	mean(—$\hat{\beta}_{10} - \hat{\beta}_{15}$—)	$\sqrt{\text{mean}(\hat{\beta}_{10} - \hat{\beta}_{15})^2}$
200	0.5	0.0067	0.0228	2	0.0114	0.0348
500	0.5	0.0086	0.0179	2	0.0118	0.0253
1000	0.5	0.0079	0.0130	2	0.0120	0.0204

Table 5: Compare Empirical std and Theorectical std

		Mixed Normal						Laplace					
		AIC			BIC			AIC			BIC		
		empirical	theo	bias	empirical	theo	bias	empirical	theo	bias	empirical	theo	bias
200	λ	0.0654	0.0502	-23.2%	0.0664	0.0487	-26.7%	0.1502	0.1053	-29.9%	0.1420	0.1243	-12.5%
	β	0.1161	0.0978	-15.8%	0.1215	0.0964	-20.7%	0.3466	0.2479	-28.5%	0.3149	0.2923	-7.2%
500	λ	0.0520	0.0393	-24.4%	0.0531	0.0378	-28.8%	0.1113	0.0850	-23.6%	0.1042	0.0952	-8.6%
	β	0.0768	0.0670	-12.8%	0.0795	0.0663	-16.4%	0.2117	0.1786	-15.6%	0.2024	0.1960	-3.2%
1000	λ	0.0333	0.0246	-26.1%	0.0338	0.0237	-29.9%	0.0671	0.0575	-14.3%	0.0634	0.0606	-4.4%
	β	0.0544	0.0477	-12.3%	0.0561	0.0473	-15.7%	0.1526	0.1278	-16.3%	0.1471	0.1359	-7.6%

Table 6: Estimates of School District Tax in Iowa

	Adjacency			County		
	PMLE	AIC	BIC	PMLE	AIC	BIC
λ	0.2535	0.1757	0.1589	0.1218	-0.0014	0.0145
	(0.0895)	(0.0340)	(0.0479)	(0.0640)	(0.0255)	(0.0314)
Income ($1000)	-0.3372	-0.2561	-0.2415	-0.3536	-0.2356	-0.2098
	(0.0658)	(0.0229)	(0.0336)	(0.0659)	(0.0228)	(0.0377)
White (%)	0.0672	0.0277	0.0113	0.0658	0.0262	0.0021
	(0.0279)	(0.0101)	(0.0128)	(0.0281)	(0.0077)	(0.0129)
State aid/pupil ($100)	-0.0500	0.1831	0.1324	-0.0235	-0.1307	0.2060
	(0.2445)	(0.1099)	(0.0934)	(0.2471)	(0.0588)	(0.0793)
Pupil/taxpayer (%)	1.3393	1.1503	1.3550	1.3322	0.6879	1.2754
	(0.5648)	(0.2002)	(0.2930)	(0.5706)	(0.1818)	(0.2703)
Property rate (per thousand)	-0.4635	-0.2919	-0.3165	-0.4652	-0.2270	-0.3677
	(0.1317)	(0.0414)	(0.0545)	(0.1328)	(0.0357)	(0.0554)
Over 65 (%)	-0.0227	-0.0910	-0.0201	-0.0052	-0.0360	0.0032
	(0.0744)	(0.0369)	(0.0368)	(0.0746)	(0.0269)	(0.0334)
College graduates (%)	-0.0058	0.0188	-0.0171	-0.0048	0.0015	-0.0317
	(0.0513)	(0.0129)	(0.0236)	(0.0520)	(0.0156)	(0.0256)
Constant	13.7846	-	-	15.0122	-	-
	(4.7004)	-	-	(4.7037)	-	-
σ	4.7704	-	-	4.8103	-	-
AIC	1917.9	1364.4	-	1922.0	1369.9	-
BIC	1956.8	-	1427.4	1960.9	-	1439.3
Number of sieves	-	11	7	-	11	8
Sample Size	361					

PMLE: parametric MLE, i.e., estimate the model under the normal distribution assumption of $\epsilon_{i,N}$

Table 7: Estimates of School District Tax in Iowa

	Adjacency			County	
	PMLE	AIC	BIC	PMLE	AIC/BIC
λ	0.2541	0.1744	0.1334	0.1224	0.0726
	(0.0893)	(0.0384)	(0.0558)	(0.0636)	(0.0301)
Income ($1000)	-0.3417	-0.2321	-0.2814	-0.3572	-0.2610
	(0.0526)	(0.0179)	(0.0499)	(0.0524)	(0.0217)
White (%)	0.0676	0.0357	0.0424	0.0661	0.0104
	(0.0277)	(0.0160)	(0.0491)	(0.0279)	(0.0125)
State aid/pupil ($100)	-0.0436	0.0341	-0.0343	-0.0181	0.1541
	(0.2378)	(0.1042)	(0.1435)	(0.2400)	(0.0846)
Pupil/taxpayer (%)	1.3547	0.9250	1.2077	1.3452	1.2511
	(0.5481)	(0.2607)	(0.3666)	(0.5528)	(0.2377)
Property rate (per thousand)	-0.4652	-0.2789	-0.3207	-0.4666	-0.3194
	(0.1309)	(0.0441)	(0.0932)	(0.1319)	(0.0470)
Over 65 (%)	-0.0211	-0.0993	-0.0388	-0.0039	-0.0080
	(0.0731)	(0.0380)	(0.0656)	(0.0732)	(0.0293)
Constant	13.7137	-	-	14.9521	-
	(4.6569)	-	-	(4.6569)	-
σ	4.7704	-	-	4.8103	-
AIC	1915.9	1362.0	-	1920.0	1368.9
BIC	1950.9	-	1420.7	1955.0	1423.3
Number of sieves	-	10	7		7
Sample Size			361		

PMLE: parametric MLE, i.e., estimate the model under the normal distribution assumption of $\epsilon_{i,N}$