QML estimation of a spatial autoregressive model with endogenous heterogeneity

Yanli Lin* Yichun Song† Yang Yang‡

October 18, 2022

Abstract

This paper develops a spatial autoregressive (SAR) model with endogenous heterogeneity. The scalar spatial coefficient is allowed to be a bounded function of a spatial unit's own endogenous characteristics. Of particular interest is a structural interaction model with nonlinear and endogenous spillover effects. We propose a quasi-maximum likelihood estimation (QMLE) with a control function approach for this model, and investigate the asymptotic properties of the estimator which is verified to have good performance in finite sample Monte Carlo simulations. We apply our model to an empirical study of regional economic performance in about 1,400 subnational regions from 110 different countries and detect that the spillovers of regional productivity are heterogeneous and depend on an endogenous variable - years of education.

Keywords: Spatial autoregressive model, Endogenous continuous heterogeneity, QMLE

JEL classification codes: C31, C51

*Department of Economics, The Ohio State University, Columbus OH 43210-1120. Email: lin.3019@osu.edu.

†Department of Economics, The Ohio State University, Columbus OH 43210-1120. Email: song.1399@osu.edu.

‡Ma Yinchu School of Economics, Tianjin University, Tianjin 300072, China. Email: yangyang_maxim@tju.edu.cn.

1 Introduction

The spatial autoregressive (SAR) model is extensively studied in spatial econometrics and has became a standard methodological toolkit of applied researchers to deal with spatial data and social networks. However, the traditional SAR model $Y_n = \rho W_n Y_n + X_n \beta + u_n$ is quite restrictive due to the exogeneity assumptions imposed on the spatial coefficient ρ and the spatial weight matrix W_n. Also, since ρ is a scalar, the term $\rho W_n Y_n$ can only capture linear spillover effect, which is homogeneous with degree of ρ (multiplied by the "spatial lag" $W_n Y_n$) for all spatial units.

Many researchers try to extend the traditional SAR model to meet broader empirical needs by relaxing the restrictions on either the spatial weight matrix or the spatial coefficient. By considering a spatial weight matrix constructed by endogenous social/economic variables, the exogenous spatial weight assumption might be violated, on the one hand, papers such as Qu and Lee (2015), Qu et al. (2017), Qu et al. (2021) provide specifications and estimation methods for a SAR model with an endogenous weight matrix, while the spatial coefficient is still a scalar, implying a homogeneous and linear spillover effect. On the other hand, some studies, for instance Malikov and Sun (2017), consider nonlinearity and parameter heterogeneity in models of spatial dependence by allowing the parameters of a linear regression to be unspecified functions of contextual variables and employ semi-parametric or non-parametric estimators. However, the contextual variables are assumed to be exogenous, ignoring the potential endogeneity concern.

The major contribution of this paper is that we develop an endogenous heterogeneous coefficient SAR (EHSAR) model, relaxing both the linear and exogenous assumptions on the spatial interaction term, and propose a likelihood estimation approach for the model. We assume that the spatial coefficient is heterogeneous, which is a bounded known function of a spatial unit's own characteristics (combined with their associated parameters). Those characteristics are likely to be correlated with the final outcome. The new model is able to capture heterogeneous spillover effects for all spatial units, similar to Malikov and Sun (2017), but allowing the spatial interactions (via the underlying contextual variables) to be endogenous. Although assuming a known function with parameters for the spatial coefficient seems not as flexible as the unknown smooth function settings in Malikov and Sun (2017), the estimated results are more clear and easier for interpretation. That is, using our setting, we can identify the sources of heterogeneity and measure their corresponding magnitudes by the significance levels and values of the parameters inside and outside the known function, while it's a harder task for semi-parametric or non-parametric estimators for unknown functions. To estimate the EHSAR model, we incorporate the control function approach with the likelihood estimation method and discuss the properties of the QML estimator. The likelihood approach enables us to compare different model specifications and select the best candidate according to likelihood based benchmarks, for instance the Akaike weight.

Our EHSAR model can be applied to various fields in economics. In urban and regional economics, the intensity of housing price spillovers may vary with regional economic growth and population mobility. In labor economics, researchers may be interested in the degree of peer effects associated with family income in educational outcomes. In public economics, regarding tax/expenditure competition among local governments, decision makers might take the labor marker condition and fiscal constraint into account, resulting in differing spillover effects of the economic outcomes. For the aforementioned examples, our proposed model would be a useful tool since the spatial interactions might depend on a spatial unit's own characteristics, which are potentially endogenous. In this paper, we apply the model to empirically study spatial spillovers of regional productivity and identify years of education as an endogenous heterogeneity source for the spatial dependence in 110 different countries.

The rest of the paper proceeds as follows. Section 2 introduces the model specification, provides its economic foundation, discusses the parameter space and function form for the heterogeneity component, and demonstrates the QMLE approach. Section 3 specifies the spatial topological structure and establishes the asymptotic properties of the estimator, the finite sample performance of which is investigated via Monte Carlo (MC) experiments in Section 4. Section 5 presents an empirical application on regional productivity spillovers and Section 6 concludes the paper. The QMLE derivation and detailed mathematical proofs are in Appendix A and B.

2 The EHSAR model

2.1 Model specification

We consider the following cross-sectional endogenous heterogeneous coefficient SAR (EHSAR) model

$$y_{i,n} = \psi(\Lambda, z_{i,n}) \sum_{j \neq i} w_{ij,n} y_{j,n} + x'_{1,in}\beta + v_{i,n} \tag{1}$$

where $i = 1, ..., n$ is the spatial unit, $y_{i,n}$ is the outcome at the location i, $w_{ij,n}$ measures the relative strength of linkage between locations $i \neq j$ ($w_{ii,n}$ for all i), $x_{1,in}$ is a k_1-dimensional vector of observed exogenous regressors at location i, $v_{i,n}$ is a scalar error term. β is a k_1-dimensional vector of parameters. The endogenous heterogeneity in spatial spillovers is assumed to be captured by $\psi(\Lambda, z_{i,n})$, which is a known (scalar) parameter function of $z_{i,n}$, an h-dimensional vector of observed endogenous variables, and $\Lambda = (\lambda_1, \cdots, \lambda_P)'$, a P-dimensional vector of parameters. Denote $Y_n = (y_{1,n}, \cdots, y_{n,n})'$, $V_n = (v_{1,n}, \cdots, v_{n,n})'$, the $n \times n$ matrix $W_n = (w_{ij,n})$, the $n \times k_1$ matrix $X_{1n} = (x_{1,1n}, \cdots, x_{1,nn})'$, and the $n \times h$ matrix $Z_n = (z_{1,n}, \cdots, z_{n,n})'$. Then we can write the EHSAR model in the matrix form as

$$Y_n = \Psi(\Lambda, Z_n) W_n Y_n + X_{1n}\beta + V_n \tag{2}$$

where $\Psi(\Lambda, Z_n) = \begin{bmatrix} \psi(\Lambda, z_{1,n}) & & \\ & \ddots & \\ & & \psi(\Lambda, z_{n,n}) \end{bmatrix} \equiv \text{diag}(\psi(\Lambda, z_{i,n}))$ is an $n \times n$ diagonal matrix of spatial parameter functions.

In this paper, we model the endogenous variables $z_{i,n}$ by the regression equation

$$z_{i,n} = \Gamma' x_{2,in} + \varepsilon_{i,n} \tag{3}$$

where $x_{2,in}$ is a k_2-dimensional vector of observed exogenous regressors, Γ is a $k_2 \times h$ matrix of coefficients, $\varepsilon_{i,n}$ is an h-dimensional vector of error terms. Denote the $n \times k_2$ matrix $X_{2n} = (x_{2,1n}, \cdots, x_{2,nn})'$ and the $n \times h$ matrix $\varepsilon_n = (\varepsilon_{1,n}, \cdots, \varepsilon_{n,n})'$. Equation (3) has a matrix form

$$Z_n = X_{2n}\Gamma + \varepsilon_n. \tag{4}$$

We model the endogeneity of the heterogeneity component $\Psi(\Lambda, Z_n)$ from the correlation between $v_{i,n}$ and $\varepsilon_{i,n}$ for all i, below is the formal assumption.

Assumption 1. *The error terms $v_{i,n}$ and $\varepsilon_{i,n}$ have a joint distribution: $\left(v_{i,n}, \varepsilon'_{i,n}\right)' \overset{i.i.d.}{\sim} (0, \Sigma_{v\varepsilon})$, where $\Sigma_{v\varepsilon} = \begin{bmatrix} \sigma_v^2 & \sigma'_{v\varepsilon} \\ \sigma_{v\varepsilon} & \Sigma_\varepsilon \end{bmatrix}$ is positive definite, σ_v^2 is a scalar variance, covariance $\sigma_{v\varepsilon} = (\sigma_{v\varepsilon_1}, \cdots, \sigma_{v\varepsilon_h})'$ is a h-dimensional vector, and Σ_ε is a $h \times h$ matrix. The $\sup_{i,n} E\left|v_{i,n}\right|^{4+\delta_\varepsilon}$ and $\sup_{i,n} E\left\|\varepsilon_{i,n}\right\|^{4+\delta_\varepsilon}$ exist for some $\delta_\varepsilon > 0$. Furthermore, $E(v_{i,n} \mid \varepsilon_{i,n}) = \varepsilon'_{i,n}\delta$ and $Var(v_{i,n} \mid \varepsilon_{i,n}) = \sigma_\xi^2$.*

Assumption 1 is also imposed in Qu and Lee (2015). It is general because specific distribution on disturbances is not required. From the two conditional moments assumptions, we have the h-dimensional vector $\delta = \Sigma_\varepsilon^{-1}\sigma_{v\varepsilon}$ and the scalar $\sigma_\xi^2 = \sigma_v^2 - \sigma'_{v\varepsilon}\Sigma_\varepsilon^{-1}\sigma_{v\varepsilon}$. Denote $\xi_n = V_n - \varepsilon_n\delta$, its mean and variance conditional on ε_n would be zero and $\sigma_\xi^2 I_n$ respectively. The EHSAR model (2) becomes

$$Y_n = \Psi(\Lambda, Z_n) W_n Y_n + X_{1n}\beta + (Z_n - X_{2n}\Gamma)\delta + \xi_n \tag{5}$$

where $(Z_n - X_{2n}\Gamma)$ are added as control variables to control the endogeneity of $\Psi(\Lambda, Z_n)$.

The traditional SAR model has a homogeneous (scalar) spatial coefficient for all spatial units. Our EHSAR model make a generalization to accommodate spatial parameter heterogeneity by assuming that the conventional spatial coefficient can be varied with a nonlinear transformation of some contextual variables (and the associated parameters) of a spatial unit, where those contextual variables can co-move with the outcome. Studies of the spillover effects of tax and trade policy among different countries provide another example of applications. Factors, such as population composition, unemployment rate, inflation, and so on, are determinants for government policies, thus enabling nonlinearities[1] of those (potential endogenous) contextual variables in the heterogeneity component from the spatial interaction term helps to obtain realistic variations of spillover effects that might depend on those underlying factors. Ignoring this feature by employing the traditional SAR model might lead to model mis-specification or misleading empirical conclusions. In this sense, the EHSAR model is a flexible alternative to the traditional SAR model.

In current setting, the heterogeneity component is assumed to be a nonlinear function of the observed (endogenous) characteristic at location i. One might also consider introducing location j's observed characteristics $z_{j,n}$ into the nonlinear transformation part as additional endogeneity sources since some spatial units may have strategic actions based on other locations' information. Although it seems an interesting idea, there are two major concerns which prevent us from adding $z_{j,n}$ into the known function: first, a heterogeneity component with $z_{j,n}$ may cause identification problem due to high dimensions of the parameters; second, with such setting, it's not easy to separately identify the spillover effects and neighborhood effects (externality), or the peer and contextual effects in the labor economics. Adding a spatial Durbin term as new regressor would be more appropriate to capture the neighborhood effects for the second concern. Regarding the first concern, the alternative specification would be forming an endogenous spatial weight matrix constructed by "economic distance" in Qu and Lee (2015), for instance $1/|z_{i,n} - z_{j,n}|$, thus incorporating the information from location j without worrying about increasing the number of parameters. A much richer EHSAR model may combine our current setting with their endogenous spatial weight matrix specification, using the control function approach to control the endogeneity of both $\Psi(\Lambda, Z_n)$ and W_n, the QML estimator (and its asymptotic properties) presented below can be used for estimation with slight modifications.

1. As noted in Paelinck and Klaassen (1979), highly nonlinear specifications are more common than linear models for econometric relations in space.

2.2 Economic foundation, parameter space, and functional form of $\psi(\Lambda, z_{i,n})$

Similar to the traditional SAR model, the EHSAR model also has a game structure, which can be regarded as a reaction function on the Nash equilibrium of a static complete information game of players with heterogeneous believes on other players' outcomes processing with linear-quadratic utilities. Suppose there are n agents, who choose actions to maximize their utilities. Denote $y_{i,n}$ as an agent i's action, assume that the cost for $y_{i,n}$ is $\frac{1}{2}y_{i,n}^2$, and that the corresponding benefit from his/her action is proportional to $y_{i,n}$, depending on an agent i's characteristics and his/her belief on other agents' actions, i.e., $y_{i,n}\left[\psi\left(\Lambda, z_{i,n}\right)\sum_{j=1}^{n} w_{ij,n}y_{j,n} + x_{1,in}^{'}\beta + \left(z_{i,n} - \Gamma' x_{2,in}\right)'\delta + \xi_{i,n}\right]$, which can be substitute or complementary depending on the sign of $\psi\left(\Lambda, z_{i,n}\right)$. Then agent i's utility is

$$u_{i,n}\left(y_{i,n}\right) = y_{i,n}\left[\psi\left(\Lambda, z_{i,n}\right)\sum_{j=1}^{n} w_{ij,n}y_{j,n} + x_{1,in}^{'}\beta + \left(z_{i,n} - \Gamma' x_{2,in}\right)'\delta + \xi_{i,n}\right] - \frac{1}{2}y_{i,n}^2$$

where W_n, $x_{1,in}$, $x_{2,in}$, $z_{i,n}$ and $\xi_{i,n}$ are public information for all agents. An agent i maximizes utility with respect to $y_{i,n}$ given $Y_{-i,n} = \left(y_{1,n}, \cdots, y_{i-1,n}, y_{i+1,n}, \cdots, y_{n,n}\right)^{'}$, $x_{1,in}$, $x_{2,in}$, $z_{i,n}$ and $\xi_{i,n}$, yielding the following best response function

$$y_{i,n}\left(Y_{-i,n}\right) = \psi\left(\Lambda, z_{i,n}\right)\sum_{j=1}^{n} w_{ij,n}y_{j,n} + x_{1,in}^{'}\beta + \left(z_{i,n} - \Gamma' x_{2,in}\right)'\delta + \xi_{i,n}, \ \ i = 1, ..., n$$

When the unique solution exists for the best response function system, equation (5) represents the unique Nash equilibrium of this game. There can be other theoretical justifications based on empirical applications of this model. For instance, the utility can be divided into the private and social components for social interactions

$$\widetilde{u}_{i,n}\left(y_{i,n}\right) = y_{i,n}\left[x_{1,in}^{'}\beta + \left(z_{i,n} - \Gamma' x_{2,in}\right)'\delta + \xi_{i,n}\right] - \frac{1}{2}\left[y_{i,n} - \psi\left(\Lambda, z_{i,n}\right)\sum_{j=1}^{n} w_{ij,n}y_{j,n}\right]^2$$

where the first component is the private utility for an agent i's action $y_{i,n}$ and the second component represents a conformity effect with his/her friends (see, Brock and Durlauf (2001)).

Denote $S_n\left(\Lambda\right) = I_n - \Psi\left(\Lambda, Z_n\right)W_n$. In order to guarantee a unique solution of the system of equations, or equivalently, the existence and uniqueness of the Nash equilibrium, it's required that $S_n^{-1}\left(\Lambda\right)$ is uniformly bounded. The EHSAR model is thus stable if $\left\|\Psi\left(\Lambda, Z_n\right)W_n\right\|_\infty < 1$, and has $Y_n = S_n^{-1}\left(\Lambda\right)\left[X_{1n}\beta + (Z_n - X_{2n}\Gamma)\delta + \xi_n\right] = \sum_{l=0}^{\infty}\left[\Psi\left(\Lambda, Z_n\right)W_n\right]^l\left[X_{1n}\beta + (Z_n - X_{2n}\Gamma)\delta + \xi_n\right]$ as the reduced form equation. $\left[\Psi\left(\Lambda, Z_n\right)W_n\right]^l$ can be regarded as the influence of the l-th layer neighborhood characteristics on each spatial unit. $\|\Psi(\Lambda, Z_n) W_n\|_\infty < 1$ implies that the impact of the l-th order contiguous neighbors decreases geometrically as l increases. Since $\Psi(\Lambda, Z_n)$ is a diagonal matrix, we have $\|\Psi(\Lambda, Z_n) W_n\|_\infty \leq \|W_n\|_\infty \cdot \sup_{\Lambda,z} |\psi(\Lambda, z_{i,n})|$. A necessary condition is that $\psi(\cdot)$ is a globally bounded function with $|\psi(\cdot)| \leq \frac{1}{\|W_n\|_\infty}$. This requirement rules out linear functions, such as $\psi(\Lambda, z_{i,n}) = \lambda_0 + \lambda_1 z_{1,in} + \cdots + \lambda_h z_{h,in}$, due to its unboundedness. Also, since $\psi(\Lambda, z_{i,n})$ is part of the spatial interaction term, which captures the heterogeneous spatial correlations, intuitively, there should be no spatial correlation when $\Lambda = 0$, thus it's reasonable to assume that $\psi(0, z_{i,n}) \equiv 0$ for all i. Based on these two requirements, transformations of linear functions would be good candidates for $\psi(\cdot)$. A useful class of functions is $\rho F(\lambda_0 + \lambda_1 z_{1,in} + \cdots + \lambda_h z_{h,in})$ with $F(\cdot)$ being a continuous probability distribution function on $\mathbb{R}$. With this form, ρ can capture the magnitude and direction of the spillover effect, and $F(\cdot)$ measures the degree of heterogeneity. For example, $F(\cdot)$ can be the distribution function of standard normal distribution $\Phi(\cdot)$, or the logistic function $1/[1 + \exp\{-(\lambda_0 + \lambda_1 z_{1,in} + \cdots + \lambda_h z_{h,in})\}]$. In fact, based on specific empirical needs, the linear component $(\lambda_0 + \lambda_1 z_{1,in} + \cdots + \lambda_h z_{h,in})$ can be replaced by other unbounded functions , for instance, quadratic functions, exponential functions, linear functions with cross terms, higher order polynomial functions, etc.

2.3 The quasi-maximum likelihood estimation

As in White (1982), based on Assumption 1, we can write down the log quasi-likelihood function under a normal distributional specification as:

$$\begin{aligned} \ln L_n = &-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma_{v\varepsilon}| + \ln|S_n(\Lambda)| \\ &-\frac{1}{2}\left[(S_n(\Lambda)Y_n - X_{1n}\beta)', (\mathrm{vec}(Z_n - X_{2n}\Gamma))'\right]\left(\Sigma_{v\varepsilon}^{-1}\otimes I_n\right)\begin{bmatrix} S_n(\Lambda)Y_n - X_{1n}\beta \\ \mathrm{vec}(Z_n - X_{2n}\Gamma)\end{bmatrix} \end{aligned} \quad (6)$$

By the partitioned quadratic formulation that

$$\left(v_{i,n}, \varepsilon_{i,n}'\right)\Sigma_{v\varepsilon}^{-1}\left(v_{i,n}, \varepsilon_{i,n}'\right)' = \left(v_{i,n} - \sigma_{v\varepsilon}'\Sigma_\varepsilon^{-1}\varepsilon_{i,n}\right)'\left(\sigma_v^2 - \sigma_{v\varepsilon}'\Sigma_\varepsilon^{-1}\sigma_{v\varepsilon}\right)^{-1}\left(v_{i,n} - \sigma_{v\varepsilon}'\Sigma_\varepsilon^{-1}\varepsilon_{i,n}\right) + \varepsilon_{i,n}'\Sigma_\varepsilon^{-1}\varepsilon_{i,n}, \quad (7)$$

the above log quasi-likelihood function can be simplified

$$
\begin{aligned}
\ln L_n(\theta) = & -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_\xi^2 + \ln|S_n(\Lambda)| - \frac{n}{2}\ln|\Sigma_\varepsilon| - \frac{1}{2}\sum_{i=1}^{n}\left(z_{i,n}' - x_{2,in}'\Gamma\right)\Sigma_\varepsilon^{-1}\left(z_{i,n} - \Gamma' x_{2,in}\right) \\
& - \frac{1}{2\sigma_\xi^2}\left[S_n(\Lambda)Y_n - X_{1n}\beta - (Z_n - X_{2n}\Gamma)\delta\right]'\left[S_n(\Lambda)Y_n - X_{1n}\beta - (Z_n - X_{2n}\Gamma)\delta\right]
\end{aligned}
\tag{8}
$$

where $\theta = \left(\Lambda', \beta', \mathrm{vec}(\Gamma)', \sigma_\xi^2, \alpha', \delta'\right)'$ with α being a $J-$dimensional column vector of distinct elements in Σ_ε. The QMLE $\hat{\theta} = \arg\max_{\theta\in\Theta}\ln L_n(\theta)$.

When the vector of parameters θ has high dimensions, direct optimization using the likelihood function (8) might be computationally demanding. So we can concentrate out some parameters before numerical computations. Denote $\xi_n(\theta) = S_n(\Lambda)Y_n - X_{1n}\beta - (Z_n - X_{2n}\Gamma)\delta$ and $\varepsilon_n(\Gamma) = Z_n - X_{2n}\Gamma$, from the first order derivatives of the log quasi-likelihood function (8) in Appendix A, setting $\frac{\partial \ln L_n(\theta)}{\partial\beta} = 0$, $\frac{\partial \ln L_n(\theta)}{\partial\sigma_\xi^2} = 0$ and $\frac{\partial \ln L_n(\theta)}{\partial\delta} = 0$ yields

$$
\beta = (X_{1n}'X_{1n})^{-1}X_{1n}'[S_n(\Lambda)Y_n - \varepsilon_n(\Gamma)\delta];\ \sigma_\xi^2 = \frac{1}{n}\xi_n'(\theta)\xi_n(\theta);\ \delta = [\varepsilon_n'(\Gamma)\varepsilon_n(\Gamma)]^{-1}\varepsilon_n'(\Gamma)[S_n(\Lambda)Y_n - X_{1n}\beta].
$$

Combining these results, we can represent β, σ_ξ^2 and δ by the remaining parameters $\widetilde{\theta} = (\Lambda', \mathrm{vec}(\Gamma)', \alpha')'$

$$
\begin{aligned}
\beta(\widetilde{\theta}) = & \left\{I_{k_1} - (X_{1n}'X_{1n})^{-1}X_{1n}'\varepsilon_n(\Gamma)[\varepsilon_n'(\Gamma)\varepsilon_n(\Gamma)]^{-1}\varepsilon_n'(\Gamma)X_{1n}\right\}^{-1} \\
& \cdot (X_{1n}'X_{1n})^{-1}X_{1n}'\left\{I_n - \varepsilon_n(\Gamma)[\varepsilon_n'(\Gamma)\varepsilon_n(\Gamma)]^{-1}\varepsilon_n'(\Gamma)\right\}S_n(\Lambda)Y_n; \\
\delta(\widetilde{\theta}) = & \left\{I_h - [\varepsilon_n'(\Gamma)\varepsilon_n(\Gamma)]^{-1}\varepsilon_n'(\Gamma)X_{1n}(X_{1n}'X_{1n})^{-1}X_{1n}'\varepsilon_n(\Gamma)\right\}^{-1} \\
& \cdot [\varepsilon_n'(\Gamma)\varepsilon_n(\Gamma)]^{-1}\varepsilon_n'(\Gamma)[I_n - X_{1n}(X_{1n}'X_{1n})^{-1}X_{1n}']S_n(\Lambda)Y_n; \\
\sigma_\xi^2(\widetilde{\theta}) = & \frac{1}{n}\xi_n'(\widetilde{\theta})\xi_n(\widetilde{\theta}),
\end{aligned}
$$

where $\xi_n(\widetilde{\theta}) = S_n(\Lambda)Y_n - X_{1n}\beta(\widetilde{\theta}) - (Z_n - X_{2n}\Gamma)\delta(\widetilde{\theta})$. Then, we can obtain the following concentrated log quasi-likelihood function

$$
\begin{aligned}
Q_n(\widetilde{\theta}) = & -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_\xi^2(\widetilde{\theta}) + \ln|S_n(\Lambda)| - \frac{n}{2}\ln|\Sigma_\varepsilon| - \frac{1}{2}\sum_{i=1}^{n}(z_{i,n}' - x_{2,in}'\Gamma)\Sigma_\varepsilon^{-1}(z_{i,n} - \Gamma' x_{2,in}) \\
& - \frac{1}{2\sigma_\xi^2(\widetilde{\theta})}[S_n(\Lambda)Y_n - X_{1n}\beta(\widetilde{\theta}) - (Z_n - X_{2n}\Gamma)\delta(\widetilde{\theta})]'[S_n(\Lambda)Y_n - X_{1n}\beta(\widetilde{\theta}) - (Z_n - X_{2n}\Gamma)\delta(\widetilde{\theta})]
\end{aligned}
\tag{9}
$$

Since there are $(k_1 + h + 1)$ less parameters in $Q_n(\widetilde{\theta})$ than those in $\ln L_n(\theta)$, computation time can be saved during the maximization algorithm.

3 Asymptotic properties

In this section, we establish consistency and asymptotic normality of the QMLE. Due to the nonlinear structure in the heterogeneity component of our EHSAR model, we build the topological specification for a cross-section unit i.

Assumption 2. *The lattice $D \subset \mathbb{R}^{d_0}$, $d_0 \geq 1$, is infinitely countable. The location $l : \{1, \cdots, n\} \rightarrow D_n \subset D$ is a mapping of unit i to its location $l(i) \in D_n \subset \mathbb{R}^{d_0}$. All elements in D are located at distances of at least $\rho_0 > 0$ from each other, i.e., $\forall l(i),\ l(j) \in D : \rho_{ij} \geq \rho_0$, where ρ_{ij} is the distance between units i and j (locations $l(i)$ and $l(j)$); w.l.o.g. we assume that $\rho_0 = 1$.*

By Jenish and Prucha (2009, 2012), this setting was introduced for asymptotic inference under near-epoch dependent (NED) processes, which indicates that we employ the increasing domain asymptotics. The distance $\rho_{ij} = \|l(i) - l(j)\|_\infty$ by maximum metric. The minimum distance ρ_0 leads to avoiding extreme influence between spatial units. Below are additional assumptions for asymptotic analyses.

Assumption 3. *(i) For any i and j, the spatial weight $w_{ij,n} \geq 0$, $w_{ii,n} = 0$, and $\sup_n \|W_n\|_\infty = c_w < \infty$.*
(ii) $\psi(\Lambda, z_{i,n})$ is a globally bounded function with $\sup_{\Lambda,z} |\psi(\Lambda, z_{i,n})| = b_\psi \leq \frac{1}{c_w}$ and $\psi(0, z_{i,n}) \equiv 0$ for $\forall z$. Furthermore, $\psi(\Lambda, z_{i,n})$ is smooth and strict monotonic for each element in Λ and z given all other elements nonzero.
(iii) The parameter $\theta = \left(\Lambda', \beta', \mathrm{vec}(\Gamma)', \sigma_\xi^2, \alpha', \delta'\right)'$ is in a compact set Θ in the Euclidean space $\mathbb{R}^{k_\theta}$. Here $k_\theta = 1 + P + k_1 + k_2 h + h + J$, where P is the dimension of Λ, k_1 is the dimension of β, h is the dimension of $\sigma_{v\varepsilon}$, $k_2 h$ is the number of parameters in Γ, and J is the dimension of α with α being the vector of all distinct elements in Σ_ε. The true parameter θ_0 belongs to the interior of Θ.
(iv) Let $k \times n$ matrix X_n collect all distinct column vectors in X_{1n} and X_{2n}. All elements in X_n are deterministic and bounded in absolute value. $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ exists and is nonsingular.

Assumption 4. *The spatial weight $w_{ij,n}$ satisfy $0 \leq w_{ij,n} \leq c_1 \rho_{ij}^{-c_2 d_0}$, where $c_1 \geq 0$ and $c_2 > 3$ are constants. Furthermore, the number of columns of W_n that the column sum exceeds c_w is less than or equal to some fixed natural number that does not depend on n, denoted as N.*

Assumptions 3(*i*) and 3(*iii*) are standard in the spatial econometrics literature. Assumption 3(*ii*) is introduced to make sure that the system of EHSAR equations has exactly one solution as discussed in Section 2.2. As $\psi(\Lambda, z_{i,n})$ is strictly increasing, its derivatives exist almost everywhere. Assumption 3(*iv*) can also be found in Qu and Lee (2015), which allows one special case that X_{1n} and X_{2n} are the same when Z_n enters the heterogeneity component nonlinearly. Assumption 4 implies that when two spatial units were close to each other, the spatial weight would be larger, and that spatial units far apart are allowed to be correlated with each other, but the spatial weight should decrease sufficiently fast at a certain rate as the spatial distance ρ_{ij} increases. This assumption accommodates a situation that $w_{ij,n} = 0$ if ρ_{ij} exceeds some threshold. Besides, the additional condition that the cardinal number $|\{j : \sum_{i=1}^{n} w_{ij,n} > c_w\}| \leq N$, which allows the existence of some "stars", i.e., larger spatial units can have great impacts on other spatial units even when n increases. In order to study the asymptotic properties of the QMLE, some moment and NED properties are needed by applying the result of Propositions 1 in Jenish and Prucha (2012), which are summarized in the following proposition.

Proposition 1. *(i) Under Assumption 3, if* $\sup_{1\leq\iota\leq k_1,i,n} E|x_{1,\iota,in}|^{\eta} < \infty$, $\sup_{i,n} E\|\varepsilon_{i,n}\|^{\eta} < \infty$ *and* $\sup_{i,n} E|\xi_{i,n}|^{\eta} < \infty$ *for some* $\eta \geq 1$*, then* $\{y_{i,n}\}_{i=1}^{n}$ *and* $\{\psi(\Lambda_0, z_{i,n}) w_{i\cdot,n} Y_n\}_{i=1}^{n}$ *are uniformly* L_η *bounded.*

(ii) Under Assumptions 1-4, $\{y_{i,n}\}_{i=1}^{n}$ *is* L_2*-NED on* $\{x_{i,n}, \varepsilon_{i,n}, \xi_{i,n}\}_{i=1}^{n}$*:* $\|y_{i,n} - E[y_{i,n}|\mathcal{F}_{i,n}(s)]\|_2 \leq Cs^{(1-c_2)d_0}$ *for some constant* $C > 0$*. The same conclusion also holds for* $\{\psi(\Lambda_0, z_{i,n}) w_{i\cdot,n} Y_n\}_{i=1}^{n}$.

Assumption 5. $S_n(\Lambda)' S_n(\Lambda)$ *is not proportional to* $S_n' S_n$ *with probability one whenever* $\Lambda \neq \Lambda_0$ *where* $S_n = S_n(\Lambda_0)$.

Lemma 1. *Under Assumptions 1-5, the true parameter* θ_0 *can be identified when the elements of* Λ *are nonzero.*

Assumption 5 is an identification condition for the EHSAR model. By Rothenberg (1971), identification for QMLE with a finite sample is equivalent to $P(\ln L_n(\theta_0) \neq \ln L_n(\theta_1)) > 0$ for any $\theta_1 \neq \theta_0$, a sufficient identification result is summarized in Lemma 1. Given the general function form of $\psi(\cdot)$, the requirement for nonzero elements of Λ might seem restrictive. However, the identification result also holds when some elements of Λ are zeros for particular constructions of $\psi(\cdot)$. One example is $\psi(\Lambda, z_{i,n}) = \rho F(\varrho_0 + \varrho_1 z_{1,in} + \cdots + \varrho_h z_{h,in})$, where $F(\cdot)$ is a continuous probability distribution function on $\mathbb{R}$. When $\rho_0 \neq 0$ and some $\varrho_{\iota,0}$'s ($\iota = 0, ..., h$) are zeros, we can still identify $\Lambda = (\rho_0, \varrho_0, ..., \varrho_h)'$ given the strict monotonicity of $\psi(\Lambda, z_{i,n})$ for each element in Λ (Assumption 3*(ii)*). We need to strengthen the identification information inequality in Assumption 6 to the limit in order to show the consistency of the QMLE.

Assumption 6. $\lim\sup_{n\to\infty}[E \ln L_n(\theta) - E \ln L_n(\theta_0)] < 0$ *for any* $\theta \neq \theta_0$.

Assumption 7. *(i)* $\{\varepsilon_{i,n}, \xi_{i,n}, x_{i,n}\}_{i=1}^{n}$ *is an* α*-mixing random field with* α*-mixing coefficient* $\alpha(\mu, \nu, \zeta) \leq$

$(\mu+\nu)^{\tau}\hat{\alpha}(\zeta)$ *for some* $\tau \geq 0$*, where* $\hat{\alpha}(\zeta)$ *satisfies* $\sum_{\zeta=1}^{\infty} \zeta^{d_0-1}\hat{\alpha}(\zeta) < \infty$.

(ii) $\sup_{i,n}\|\varepsilon_{i,n}\|_5 < \infty$, $\sup_{i,n}\|\xi_{i,n}\|_5 < \infty$, $\sup_{i,k,n}\|x_{i,k,n}\|_5 < \infty$.

Proposition 2. *Under Assumptions 1-4 and 7(i),* $\sup_{\Lambda\in\Theta}\frac{1}{n}(\ln|I_n-\Psi(\Lambda,Z_n)W_n| - E\ln|I_n-\Psi(\Lambda,Z_n)W_n|) \xrightarrow{p} 0$.

Theorem 1. *Under Assumptions 1-7, the QMLE* $\hat{\theta}$ *is a consistent estimator of* θ_0.

With the identification, we still need to show the uniform convergence of $\frac{1}{n}\sup_{\theta\in\Theta}|\ln L_n(\theta) - E\ln L_n(\theta)| \xrightarrow{p} 0$ and the uniform equicontinuity of $\lim_{n\to\infty} E(\frac{1}{n}\ln(\theta))$ to establish the the consistency of the QMLE. One key step in proving the uniform convergence of the quasi log-likelihood function is to show the uniform convergence of the log determinant term (Proposition 2), the form of which is not similar to the usual form in a traditional SAR model. By referring to Qu and Lee (2013), we derive a useful formula of the Taylor series of $\ln|I_n - \Psi(\Lambda, Z_n)W_n|$, then we check its NED property and uniform boundedness. With Assumption 7 being a regularity condition for the application of the LLN in Jenish and Prucha (2012), the consistency of the QMLE is provided in Theorem 1.

Assumption 8. *For some* $\phi > 0$*, the* α*-mixing coefficient of* $\{\varepsilon_{i,n}, \xi_{i,n}, x_{i,n}\}_{i=1}^{n}$ *in Assumption 7(i) satisfies* $\sum_{\zeta=1}^{\infty}\zeta^{d_0(\tau_*+1)}\hat{\alpha}^{\frac{\phi}{4+2\phi}}(\zeta) < \infty$*, where* $\tau_* = \phi\tau/(2+\phi)$.

Assumption 9. $\Sigma_0 = \lim_{n\to\infty}\frac{1}{n}E\left(\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}\right)$ *exists and is nonsingular.*

Theorem 2. *Under Assumptions 1-9,* $\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0, \Sigma_{QMLE})$*, where* $\Sigma_{QMLE} = (\lim_{n\to\infty}\frac{1}{n}E(\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}))^{-1} \times \lim_{n\to\infty}\frac{1}{n}E(\frac{\partial \ln L_n(\theta_0)}{\partial\theta}\frac{\partial \ln L_n(\theta_0)}{\partial\theta'})(\lim_{n\to\infty}\frac{1}{n}E(\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}))^{-1}$.

By Taylor expansion, $\sqrt{n}\left(\hat{\theta}-\theta_0\right) = \left(-\frac{1}{n}\frac{\partial^2 \ln L_n(\bar{\theta})}{\partial\theta\partial\theta'}\right)^{-1}\left(\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta}\right)$, where $\bar{\theta}$ lies between θ_0 and $\hat{\theta}$. To derive the asymptotic distribution of the QMLE, we first show the asymptotic normality of the consistent root of $\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\hat{\theta})}{\partial\theta} = 0$ by showing the sequence of scores obeys a CLT by Theorem 2 in Jenish and Prucha (2012) with the regularity conditions in Assumptions 8 and 9. Then we check $\frac{1}{n}\left|\frac{\partial^2 \ln L_n(\hat{\theta})}{\partial\theta\partial\theta'} - E\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}\right| \xrightarrow{p} 0$. The asymptotic distribution is derived in Theorem 2. When $\left(v_{i,n}, \varepsilon'_{i,n}\right)'$ is jointly normal, we have MLE instead and $\Sigma_{MLE} = -\left(\lim_{n\to\infty}\frac{1}{n}\mathrm{E}(\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'})\right)^{-1}$.

4 Monte Carlo simulation

To demonstrate the finite sample properties and robustness of the proposed QML estimator for our EHSAR model, we conduct a series of Monte Carlo experiments. Specifically, we would like to investigate the finite sample performance of our estimators under different sample sizes and with multiple endogenous variables, the precision of the estimates from nonlinear and linear models if the true model is a traditional SAR model without a nonlinear heterogeneity component, and the robustness of the QMLE if the error terms don't follow a joint normal distribution and the misspecification of the function forms of $\psi(\cdot)$. The data generating processes (DGPs) are equations 2 and 4 based on Assumption 1. A row-normalized queen-based contiguity weights matrix is employed, i.e., $W_n = W^q_{R,n}$, where $W^q_{R,n} = (w^q_{R,ij,n})$ with $w^q_{R,ij,n} = \frac{w^q_{ij,n}}{\sum_{k=1}^n w^q_{ik,n}}$. The number of sample repetitions is 1,000 for each experiment in order to obtain empirical mean, empirical standard deviation (Std), and 95% coverage probability (CP)[2]. The detailed designs for the variables, true parameter settings and endogeneity sources are illustrated below according to specific issues we examine.

Different sample sizes. Table 1 shows the performance of our estimator under different sample sizes $n = 100,\ 256,\ 361,\ 400,\ 625$. In this simulation experiment, $x_{1,in} = (x_{1,1,in}, x_{1,2,in})'$, $x_{2,in} = (x'_{1,in}, x_{2,1,in})'$, where $x_{1,1,in} = 1$ and $(x_{1,2,in}, x_{2,1,in})' \sim i.i.d.\ N(\mathbf{0}_{2\times 1}, \Sigma_{x,0})$ with $\Sigma_{x,0} = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}$.[3] $\Psi(\Lambda_0, Z_n) \equiv \text{diag}\left(\rho_0 F(z'_{i,n}\lambda_0)\right)$ with $F(\cdot)$ being the CDF for the Logistic distribution.[4] An endogenous scalar $z_{i,n}$ is generated by $z_{i,n} = \Gamma_0' x_{2,in} + \varepsilon_{i,n}$ with $\Gamma_0 = (-0.5, 0.5, 1)'$. The endogeneity of $\Psi(\Lambda_0, Z_n)$ comes from the correlation between $v_{i,n}$ and $\varepsilon_{i,n}$, i.e., we generate the bivariate normal random variables $(v_{i,n}, \varepsilon_{i,n})$ from $i.i.d.\ N(\mathbf{0}_{2\times 1}, \Sigma_{v\varepsilon,0})$, where

$$\Sigma_{v\varepsilon,0} = \begin{pmatrix} \sigma^2_{v,0} & \sigma_{v\varepsilon,0} \\ \sigma_{v\varepsilon,0} & \sigma^2_{\varepsilon,0} \end{pmatrix} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

The true parameter values of $(\beta_0, \beta_1, \rho, \lambda, \sigma_{v\varepsilon}, \sigma_v)'$ are set to be $(-1, 4, 0.8, 0.5, 0.5, 1)'$, which are exhibited in the second column of the table. As expected, the performance of our estimator improves as the sample size increases. When $n = 100$, though we have precise estimates for all other parameters, the estimated $\hat{\lambda}$ has large bias and variance at the same time. This bias quickly shrinks as the sample size enlarges. Therefore, we recommend that our estimator should be used with moderate to large sample size.

Multiple endogenous variables. Table 2 displays the simulation results when we have more than one endogenous variables. In this experiment, we consider $z_{i,n} = (z_{1,in}, z_{2,in})'$ and add an excluded instrument variable in $x_{2,in} = (x'_{1,in}, x_{2,1,in}, x_{2,2,in})'$, where $x_{1,in} = (1, x_{1,2,in})'$. The function form for $\Psi(\Lambda_0, Z_n)$ is the same as in Table 1. $(x_{1,2,in}, x_{2,1,in}, x_{2,2,in})' \sim i.i.d.\ N(\mathbf{0}_{3\times 1}, \Sigma_{x,0})$ and

2. 95% CP represents the proportion of the 95% asymptotic-distribution-based confidence intervals that contain the true parameters.

3. $x_{1,2,in}$ and $x_{2,1,in}$ are designed to allow correlation.

4. The probability density is $f(x|\mu, \sigma) = \exp(x)/[(1 + \exp(x))^2]$.

$(v_{i,n}, \varepsilon'_{i,n}) \sim i.i.d.\ N(\mathbf{0}_{3\times 1}, \Sigma_{v\varepsilon,0})$, where

$$\Sigma_{x,0} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.3 \\ 0 & 0.3 & 1 \end{pmatrix}, \quad \Sigma_{v\varepsilon,0} = \begin{pmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{pmatrix}.$$

In the set of equations for $z_{i,n}$, $\Gamma_0 = \begin{pmatrix} -0.5 & -1 & 1 & 0.5 \\ 0 & 0.7 & -1 & 2 \end{pmatrix}'$. We also investigate the misspecification errors when one uses the traditional SAR model without accounting for the heterogeneity (denoted as "SAR"), and when one allows for heterogeneity in the SAR model, but incorrectly treats $z_{i,n}$ as exogenous (denoted as "Exogeneity"). p10 to p90 refer to the quantiles of bias.[5] Results in Table 2 indicate that the QMLE performs well when the model is correctly specified (i.e., the EHSAR model), and that large bias in estimates could occur under both misspecified models.

Non-normal joint distributions for error terms. In this experiment, we consider the same design for a univariate $z_{i,n}$ as in Table 1, but the joint distribution of the error terms $(v_{i,n}, \varepsilon_{i,n})$ is not normal. We test four non-normal cases: (1) Wishart distribution [6], (2) t distribution [7], (3) Gaussian Mixture I and (4) Gaussian Mixture II [8]. Table 3 demonstrates that in general, our estimator is robust to the error term misspecifications, except for the bimodal distribution (3) Gaussian Mixture I with

5. For instance, p50 of β_1 is obtained by subtracting its true value from the estimates and then getting the median.

6. The probability density of a Wishart distribution for a random matrix X is

$$f(X, \Sigma, \nu) = \frac{|X|^{((\nu-d-1)/2)} \exp(0.5\mathrm{tr}(\Sigma^{-1}X))}{2^{\nu d/2}\pi^{(d(d-1))/4}|\Sigma|^{\nu/2}\Gamma(\nu/2)\cdots\Gamma(\nu-(d-1)/2)}.$$

In the simulation, we set $d = 2$, $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, and $\nu = 2$. For each $i \in n$, we generate a random matrix and pick the diagonal elements as $(v_{i,n}, \varepsilon_{i,n})$.

7. Multivariate t distribution for a vector $(v_{i,n}, \varepsilon_{i,n})$ has the density

$$f(x, \Sigma, \nu) = \frac{1}{|\Sigma|^{1/2}} \frac{1}{\sqrt{(\nu\pi)^2}} \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)} \left(1 + \frac{x'\Sigma^{-1}x}{\nu}\right)^{-(\nu+d)/2}$$

In the simulation, we set $d = 2$, $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ and $\nu = 3$.

8. The two cases of Gaussian Mixture distribution follow: Case (1), a mixture of normal distributions equals

$$0.4 \times N\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right) + 0.4 \times N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right) + 0.1 \times N\left(\begin{bmatrix} -4 \\ -4 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right)$$

In Case (2), we have

$$\frac{1}{3} \times N\left(\begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right) + \frac{1}{3} \times N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right) + \frac{1}{3} \times N\left(\begin{bmatrix} -1.5 \\ -1.5 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right)$$

In case (1), the mixture gives us a bimodal density, while case (2) is still kept to be unimodal.

large bias and lower 95% CPs, which is not unexpected since a bimodal density is distinct from a normal one.

Different F functions and F function misspecification. One of the drawbacks of our estimator is that the nonlinear $\psi(\cdot)$ function governing the heterogeneity in equation (2) needs to be fixed before doing estimation, constituting a potential source of misspecification. Suppose $\Psi(\Lambda_0, Z_n) \equiv \text{diag}\left(\rho_0 F\left(z_{i,n}'\lambda_0\right)\right)$ and we consider the same design for a univariate $z_{i,n}$ in Table 1, here we study the robustness towards different F functions and consequences if the F in the DGP (F_{gen}) is not the same as the F used in the estimation (F_{est}). In Table 4, the first row points to F_{gen} and the second row is F_{est}. For instance, model (3) indicates that the data is generated by $F_{gen} = N(0,1)$ and $F_{est} = t(2)$, t distribution with degree of freedom equals 2 [9]. Estimates for parameters except that for λ are robust to F function misspecification, but there is large discrepancy in $\hat{\lambda}$. Due to the different underlying F functions, $\hat{\lambda}$ can't be comparable directly. Instead, each subfigure in Figure 1 shows the distribution for the estimates of the heterogeneity component $\rho F(z_{i,n}'\lambda)$ from the 400 simulated sample points in the first repetition of this experiment[10], which is not very different to the true distribution.

The DGP is a traditional SAR model. If there is indeed no heterogeneity, i.e., $Y_n = \rho_0 W_n Y_n + X_{1n}\beta_0 + V_n$, our estimator for the EHSAR model (Table 1's design) can still capture the correct model setting, as shown in Table 5. For this specific Logistic F function, the EHSAR model accommodates it by assigning $\hat{\lambda} \approx 0$ and $\hat{\rho} \approx 2 \times 0.8$. By setting $\hat{\lambda} \approx 0$, it eliminates the heterogeneity, but leaves $F(z_{i,n}'\lambda) = 0.5$ for every sample point when F is the CDF for the Logistic distribution. Thus, the estimated $\hat{\rho}$ needs to double to write this off. Finally we end up with $\hat{\rho}F(z_{i,n}'\hat{\lambda}) \approx 0.8031$.

$z_{i,n}$ also added as regressors. In this experiment, $z_{i,n}$ is also added as a regressor (Table 1's design), which is common in practice, since the individual characteristics that influence the degree of spatial dependence may also directly determine its own outcome. Table 6 shows that our estimator naturally embodies this case and provide good estimation results.

9. Our simulation here is restricted since N(0,1), Logitstic and t are all similar unimodal distributions with the CDFs map $(-\infty, +\infty)$ to $(0, 1)$ so that the level of misspecification is not significant. Misspecification would be problematic if, for example, F_{gen} is N(0,1), but we estimate it by an Exponential CDF.

10. Note that for each experiment, we generate a total of 1000 repeated samples.

			$n=100$	$n=256$	$n=361$	$n=400$	$n=625$
β_0	-1	Mean	-1.0072	-1.0034	-1.0031	-1.0015	-1.0001
		Std	(0.1295)	(0.0774)	(0.0661)	(0.0628)	(0.0494)
		95% CP	0.9550	0.9510	0.9430	0.9500	0.9560
β_1	4	Mean	3.9967	3.9992	3.9979	3.9971	3.9973
		Std	(0.1038)	(0.0649)	(0.0547)	(0.0525)	(0.0426)
		95% CP	0.9440	0.9470	0.9590	0.9570	0.9570
ρ	0.8	Mean	0.7907	0.7959	0.7958	0.7970	0.7986
		Std	(0.1016)	(0.0633)	(0.0522)	(0.0486)	(0.0393)
		95% CP	0.9960	0.9430	0.9480	0.9500	0.9510
λ	0.5	Mean	0.5565	0.5196	0.5115	0.5094	0.5016
		Std	(0.6335)	(0.1280)	(0.1057)	(0.1001)	(0.0753)
		95% CP	0.9410	0.9520	0.9520	0.9470	0.9480
σ_{vz}	0.5	Mean	0.4991	0.5041	0.5019	0.5017	0.4998
		Std	(0.1220)	(0.0751)	(0.0651)	(0.0620)	(0.0497)
		95% CP	0.9440	0.9540	0.9490	0.9520	0.9480
σ_v	1	Mean	0.9834	0.9946	0.9954	0.9955	0.9970
		Std	(0.0727)	(0.0443)	(0.0384)	(0.0369)	(0.0299)
		95% CP	0.9420	0.9460	0.9460	0.9440	0.9440

Table 1: Simulation results under different sample sizes.

			Mean	Std	95% CP	p10	p30	p50	p70	p90
β_0	-1	SAR	-1.0242	(0.0717)	0.9340	-0.1194	-0.0594	-0.0232	0.0143	0.0693
		Exogeneity	-0.8861	(0.2463)	0.9350	-0.0224	0.0238	0.0578	0.0924	0.1547
		EHSAR	-1.0034	(0.0593)	0.9520	-0.0827	-0.0358	-0.0039	0.0304	0.0717
β_1	4	SAR	4.1979	(0.0702)	0.1960	0.1025	0.1625	0.1991	0.2345	0.2909
		Exogeneity	4.0098	(0.0784)	0.9560	-0.0656	-0.0135	0.0164	0.0473	0.0852
		EHSAR	4.0003	(0.0511)	0.9560	-0.0644	-0.0258	0.0025	0.0263	0.0657
λ_1	0.5	SAR	-	-	-	-	-	-	-	-
		Exogeneity	0.2771	(0.1020)	0.4080	-0.3489	-0.2713	-0.2211	-0.1751	-0.1017
		EHSAR	0.5129	(0.0932)	0.9380	-0.0973	-0.0389	0.0055	0.0523	0.1327
λ_2	-0.5	SAR	-	-	-	-	-	-	-	-
		Exogeneity	-0.5738	(0.1238)	0.9130	-0.2044	-0.1332	-0.0844	-0.0369	0.0544
		EHSAR	-0.5117	(0.0841)	0.9490	-0.1196	-0.0509	-0.0073	0.0364	0.0912
ρ	0.8	SAR	0.3674	(0.0338)	0.0000	-0.4769	-0.4504	-0.4331	-0.4164	-0.3874
		Exogeneity	0.9189	(0.3068)	0.9350	-0.0052	0.0200	0.0397	0.0639	0.1063
		EHSAR	0.7960	(0.0402)	0.9510	-0.0556	-0.0241	-0.0050	0.0165	0.0495
$\sigma_{v\varepsilon_1}$	0.5	SAR	-	-	-	-	-	-	-	-
		Exogeneity	-	-	-	-	-	-	-	-
		EHSAR	0.4992	(0.0597)	0.9450	-0.0767	-0.0323	-0.0000	0.0271	0.0763
$\sigma_{v\varepsilon_2}$	0.5	SAR	-	-	-	-	-	-	-	-
		Exogeneity	-	-	-	-	-	-	-	-
		EHSAR	0.4981	(0.0572)	0.9540	-0.0712	-0.0321	-0.0029	0.0258	0.0737
σ_v	1	SAR	1.2521	(0.1068)	0.3390	0.1215	0.1952	0.2485	0.2995	0.3963
		Exogeneity	0.9189	(0.3068)	0.9350	-0.2052	-0.1800	-0.1603	-0.1361	-0.0937
		EHSAR	0.9937	(0.0395)	0.9470	-0.0550	-0.0297	-0.0066	0.0146	0.0443

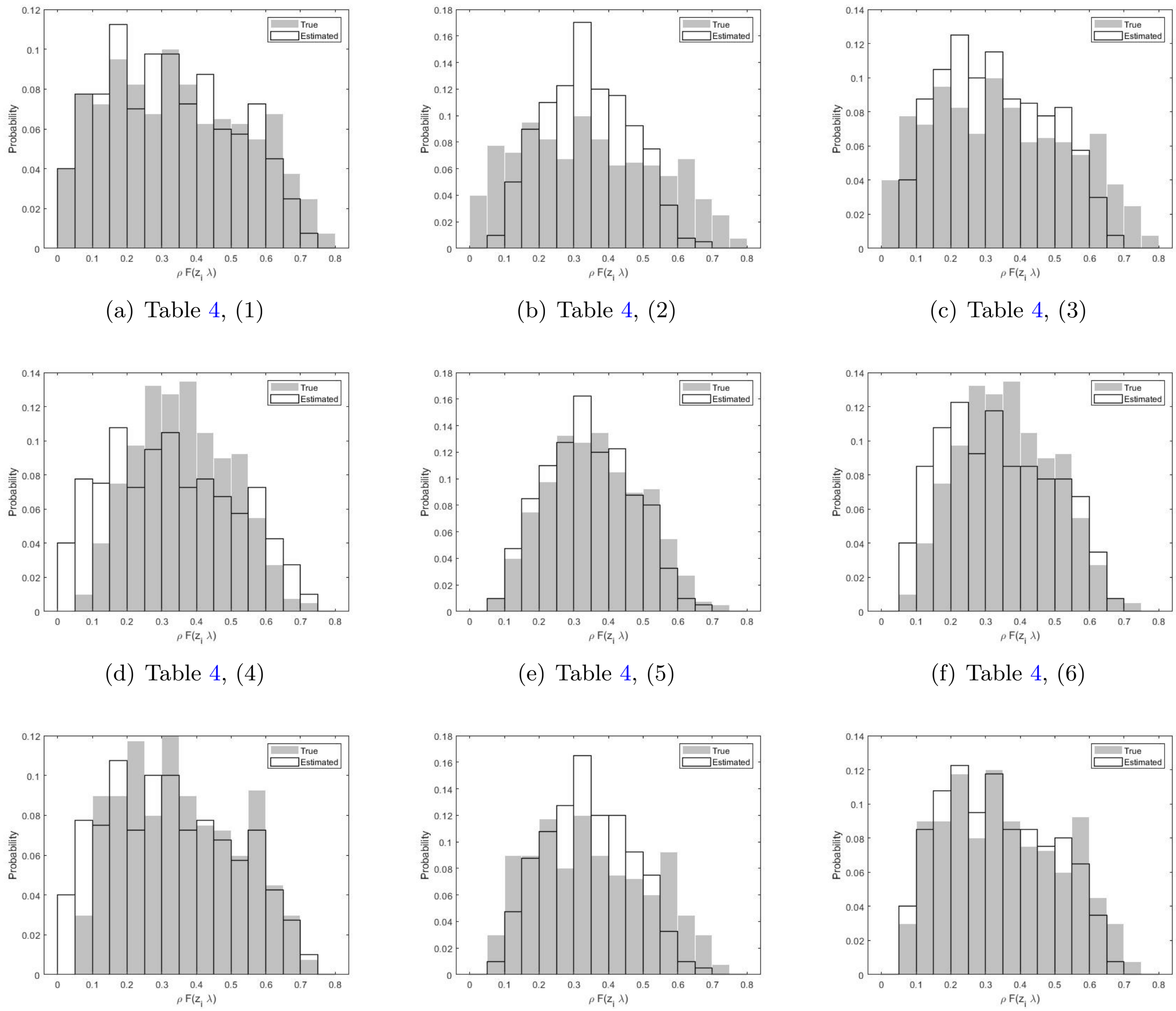
Table 2: Simulation results with multiple endogenous variables. $n = 400$.

			Mean	Std	95% CP	p10	p30	p50	p70	p90
β_0	-1	Wishart	-1.0088	(0.0757)	0.9480	-0.1003	-0.0442	-0.0046	0.0298	0.0854
		t	-1.0049	(0.1041)	0.9490	-0.1318	-0.0593	-0.0060	0.0505	0.1273
		Mixture I	-0.8875	(0.1096)	0.8260	-0.0225	0.0546	0.1121	0.1701	0.2558
		Mixture II	-0.9972	(0.0915)	0.9510	-0.1170	-0.0460	0.0030	0.0514	0.1206
β_1	4	Wishart	3.9992	(0.1059)	0.9500	-0.1359	-0.0562	-0.0027	0.0525	0.1338
		t	4.0000	(0.0945)	0.9520	-0.1240	-0.0478	-0.0010	0.0528	0.1212
		Mixture I	4.0277	(0.0998)	0.9420	-0.1019	-0.0259	0.0277	0.0808	0.1524
		Mixture II	3.9911	(0.0777)	0.9500	-0.1098	-0.0484	-0.0083	0.0275	0.0893
ρ	0.8	Wishart	0.7957	(0.1112)	0.9490	-0.1433	-0.0569	0.0029	0.0542	0.1362
		t	0.7957	(0.0780)	0.9520	-0.1058	-0.0421	-0.0008	0.0382	0.0889
		Mixture I	0.7543	(0.0714)	0.9100	-0.1353	-0.0819	-0.0482	-0.0112	0.0466
		Mixture II	0.7950	(0.0652)	0.9540	-0.0869	-0.0369	-0.0044	0.0296	0.0784
λ	0.5	Wishart	0.5753	(0.3611)	0.9860	-0.1578	-0.0531	0.0317	0.1326	0.3175
		t	0.5246	(0.2166)	0.9470	-0.2274	-0.0724	0.0188	0.1104	0.2852
		Mixture I	0.9177	(0.4042)	0.8430	-0.0694	0.1953	0.3914	0.5938	0.9022
		Mixture II	0.4809	(0.1154)	0.9530	-0.1528	-0.0829	-0.0315	0.0319	0.1362
$\sigma_{v\varepsilon}$	0.8362	Wishart	1.0105	(0.3378)	0.7150	-0.8673	-0.6477	-0.4836	-0.3040	-0.0018
	1.6092	t	1.5135	(1.4266)	0.9840	-0.6038	-0.3652	-0.1641	0.1121	0.6800
	2.2174	Mixture I	2.2878	(0.3054)	0.9520	-0.3180	-0.0850	0.0674	0.2412	0.4714
	1.4670	Mixture II	1.4744	(0.1231)	0.9470	-0.1535	-0.0565	0.0073	0.0777	0.1613
σ_v	1.8664	Wishart	1.9956	(0.1451)	0.1270	0.2689	0.3711	0.4390	0.5155	0.6413
	1.1882	t	1.6896	(0.3132)	0.9540	-0.1003	-0.0017	0.0806	0.1866	0.4080
	1.8008	Mixture I	1.8114	(0.0926)	0.9460	-0.1145	-0.0342	0.0118	0.0610	0.1262
	1.5476	Mixture II	1.5669	(0.0519)	0.9420	-0.0462	-0.0065	0.0194	0.0495	0.0835

Table 3: Simulation results under non-normally distributed error terms. $n = 400$.

			N(0,1)			Logistic			t		
			N(0,1)	Logitsitc	t(2)	N(0,1)	Logitsitc	t(2)	Normal	Logitstic	t(2)
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
β_0	-1	Mean	-1.0008	-0.9991	-0.9951	-1.0020	-1.0015	-1.0003	-1.0046	-1.0035	-1.0011
		Std	(0.0616)	(0.0614)	(0.0610)	(0.0629)	(0.0628)	(0.0626)	(0.0628)	(0.0626)	(0.0621)
		95% CP	0.9500	0.9490	0.9500	0.9480	0.9500	0.9480	0.9470	0.9480	0.9480
		p10	-0.0792	-0.0773	-0.0729	-0.0797	-0.0794	-0.07833	-0.0836	-0.0822	-0.0790
		p90	0.0793	0.0803	0.0836	0.0822	0.0822	0.08222	0.0784	0.0785	0.0797
β_1	4	Mean	3.9965	3.9947	3.9901	3.9979	3.9971	3.9953	4.0008	3.9996	3.9967
		Std	(0.0517)	(0.0516)	(0.0512)	(0.0527)	(0.0525)	(0.0522)	(0.0527)	(0.0525)	(0.0520)
		95% CP	0.9590	0.9590	0.9530	0.9570	0.9570	0.9580	0.9560	0.9590	0.9600
		p10	-0.0685	-0.0701	-0.0743	-0.0686	-0.0687	-0.0699	-0.0660	-0.0665	-0.0688
		p90	0.0648	0.0625	0.0579	0.0692	0.0680	0.06368	0.0717	0.0698	0.0657
ρ	0.8	Mean	0.7974	0.7982	0.7996	0.7969	0.7970	0.7972	0.7958	0.7963	0.7971
		Std	(0.0491)	(0.0491)	(0.0491)	(0.0486)	(0.0486)	(0.0486)	(0.0491)	(0.0490)	(0.0490)
		95% CP	0.9490	0.3680	0.9500	0.9500	0.9500	0.9500	0.9490	0.9490	0.9480
		p10	-0.0668	0.1633	-0.0633	-0.0664	-0.0657	-0.0655	-0.0678	-0.0677	-0.0660
		p90	0.0592	0.5277	0.0609	0.05776	0.0579	0.0587	0.0575	0.0578	0.0586
λ	0.5	Mean	0.5094	0.8388	0.6347	0.3125	0.5094	0.3767	0.4136	0.6795	0.5115
		Std	(0.0862)	(0.1451)	(0.1168)	(0.0602)	(0.1001)	(0.0778)	(0.0771)	(0.1287)	(0.1017)
		95% CP	0.9490	0.9490	0.8130	0.1160	0.9470	0.6230	0.7930	0.7440	0.9510
		p10	-0.0947	-0.0657	-0.0042	-0.2607	-0.1127	-0.218	-0.1805	0.0222	-0.1126
		p90	0.1223	0.0597	0.2884	-0.1116	0.1384	-0.02583	0.0122	0.3456	0.1424
σ_{vz}	0.5	Mean	0.5011	0.4990	0.4939	0.5025	0.5017	0.4996	0.5059	0.5046	0.5013
		Std	(0.0612)	(0.0611)	(0.0607)	(0.0622)	(0.0620)	(0.0617)	(0.0622)	(0.0620)	(0.0615)
		95% CP	0.9520	0.9510	0.9510	0.9490	0.9520	0.9540	0.9530	0.9510	0.9530
		p10	-0.0769	-0.0788	-0.0845	-0.0767	-0.0777	-0.0792	-0.0734	-0.0742	-0.0773
		p90	0.0790	0.0765	0.0701	0.0830	0.0822	0.0784	0.0864	0.0838	0.0801
σ_v	1	Mean	0.9952	0.9941	0.9919	0.9960	0.9955	0.9945	0.9979	0.9971	0.9953
		Std	(0.0365)	(0.0364)	(0.0362)	(0.0370)	(0.0369)	(0.0367)	(0.0369)	(0.0368)	(0.0366)
		95% CP	0.9410	0.9390	0.9430	0.9450	0.9440	0.9420	0.9470	0.9460	0.9410
		p10	-0.0525	-0.0530	-0.0551	-0.0513	-0.0517	-0.0525	-0.0503	-0.0511	-0.0525
		p90	0.0426	0.0415	0.0390	0.0439	0.0435	0.0427	0.0463	0.0452	0.0433

Table 4: Simulations results with different F functions. $n = 400$.



(a) Table 4, (1) (b) Table 4, (2) (c) Table 4, (3)

(d) Table 4, (4) (e) Table 4, (5) (f) Table 4, (6)

(g) Table 4, (7) (h) Table 4, (8) (i) Table 4, (9)

Figure 1: Distributions of the estimates for the $\rho F(z_i\lambda)$ component in models of Table 4.

			Mean	Std	95% CP	p10	p30	p50	p70	p90
		SAR	-1.0105	(0.0890)	0.9440	-0.1235	-0.0537	-0.0064	0.0396	0.0969
β_0	-1	Exogeneity	-0.9864	(0.0904)	0.9500	-0.1001	-0.0266	0.0170	0.0633	0.1215
		EHSAR	-1.0078	(0.0813)	0.9420	-0.1067	-0.0454	-0.0054	0.0335	0.0926
		SAR	3.9988	(0.0513)	0.9540	-0.0639	-0.0281	-0.0026	0.0264	0.0646
β_1	4	Exogeneity	3.8812	(0.0599)	0.4860	-0.1938	-0.1480	-0.1200	-0.0882	-0.0421
		EHSAR	3.9994	(0.0567)	0.9610	-0.0716	-0.0299	-0.0013	0.0301	0.0726
		SAR	-	-	-	-	-	-	-	-
λ	-	Exogeneity	-0.0758	(0.0176)	-	-	-	-	-	-
		EHSAR	0.0005	(0.0150)	-	-	-	-	-	-
		SAR	0.7977	(0.0140)	0.9440	-0.0199	-0.0088	-0.0019	0.0051	0.0150
ρ	0.8	Exogeneity	1.5666	(0.0300)	-	-	-	-	-	-
		EHSAR	1.5969	(0.0249)	-	-	-	-	-	-
		SAR	-	-	-	-	-	-	-	-
$\sigma_{v\varepsilon}$	0.5	Exogeneity	-	-	-	-	-	-	-	-
		EHSAR	0.5029	(0.0660)	0.9520	-0.0808	-0.0315	-0.0001	0.0348	0.0882
		SAR	0.9942	(0.0709)	0.9460	-0.0932	-0.0451	-0.0061	0.0295	0.0858
σ_v	1	Exogeneity	0.9182	(0.0666)	0.7760	-0.1682	-0.1160	-0.0849	-0.0494	0.0054
		EHSAR	0.9966	(0.0384)	0.9500	-0.0540	-0.0224	-0.0052	0.0154	0.0473

Table 5: Simulation results when the DGP is a traditional SAR model. $n = 400$.

			Mean	Std	95% CP	p10	p30	p50	p70	p90
		SAR	-0.9504	(0.1023)	0.9190	-0.0752	0.0013	0.0480	0.1052	0.1775
β_0	-1	Exogeneity	-0.8698	(0.0742)	0.5870	0.0342	0.0918	0.1290	0.1688	0.2283
		EHSAR	-1.0036	(0.0753)	0.9500	-0.0970	-0.0442	-0.0041	0.0353	0.0910
		SAR	3.7874	(0.0574)	0.0420	-0.2892	-0.2423	-0.2097	-0.1837	-0.1375
β_1	4	Exogeneity	3.7880	(0.0540)	0.024	-0.2817	-0.2399	-0.2101	-0.1850	-0.1444
		EHSAR	4.0008	(0.0666)	0.9480	-0.0817	-0.0363	-0.0009	0.0310	0.0879
		SAR	1.9980	(0.0565)	0.9530	-0.0755	-0.0311	-0.0002	0.0288	0.0673
β_z	2	Exogeneity	2.2611	(0.0466)	0.0000	0.2020	0.2378	0.2598	0.2851	0.3224
		EHSAR	1.9942	(0.0632)	0.9480	-0.0912	-0.0354	-0.0035	0.0262	0.0748
		SAR	-	-	-	-	-	-	-	-
λ	0.5	Exogeneity	0.5025	(0.0744)	0.954	-0.0873	-0.0382	-0.0022	0.0374	0.1010
		EHSAR	0.5015	(0.0691)	0.9520	-0.0817	-0.0371	0.0004	0.0342	0.0895
		SAR	0.3608	(0.0227)	0.0000	-0.4665	-0.4515	-0.4397	-0.4273	-0.4104
ρ	0.8	Exogeneity	0.7984	(0.0349)	0.0000	-0.2453	-0.2196	-0.2011	-0.1818	-0.1584
		EHSAR	0.7985	(0.0320)	0.952	-0.0438	-0.0179	-0.0009	0.0151	0.0395
		SAR	-	-	-	-	-	-	-	-
$\sigma_{v\varepsilon}$	0.5	Exogeneity	-	-	-	-	-	-	-	-
		EHSAR	0.5051	(0.0818)	0.9540	-0.0948	-0.0386	0.0025	0.0436	0.1161
		SAR	1.0102	(0.0764)	0.9430	-0.0832	-0.0319	0.0047	0.0480	0.1085
σ_v	1	Exogeneity	0.8582	(0.0600)	0.3270	-0.2180	-0.1735	-0.1436	-0.1124	-0.0665
		EHSAR	0.9975	(0.0453)	0.9440	-0.0577	-0.0262	-0.0047	0.0218	0.0574

Table 6: Simulation results when $z_{i,n}$ also enters as a regressor. $n = 400$.

5 Empirical application

To demonstrate the practical usefulness of our proposed model, we take studies on potential spatial spillovers in productivity as an example. Regional productivity is measured by income per capita in this illustration. The determinants of regional development are consistently among the most critical topics in economics. Numerous studies have provided different channels to explain the growth, including Palivos and Wang (1996), and Martin and Sunley (1998). Here we focus on the spatial analysis and check the possibility of endogenous heterogeneity in income spillovers. We combine the data set collected and made public by Gennaioli et al. (2013) (henceforth as GLLS). GLLS data contains 1,569 subnational regions[11] in 110 different countries, comprising 97% of the world's GDP.

The model we estimated is as follows,

$$\ln y_{i,n} = \rho F(z'_{i,n}\lambda) \sum_{j \neq i} w_{ij,n} \ln y_{j,n} + x'_{1,in}\beta + v_{i,n} \tag{10}$$

The dependent variable $y_{i,n}$ is the income per capita. F is a bounded Sigmoid-type function, like the Normal or Logistic CDF. The explanatory variables include, for instance, years of education, temperature, per capita oil production, population, inverse distance to the coast, and the number of ethnic groups residing in a region. The location data for all these subregions is retrieved from Sanso-Navarro et al. (2017). We then construct $w_{ij,n}$ as a row-normalized spatial weighting matrix by Rook contiguity as in Hoshino (2022) [12]. The potentially endogenous $z_{i,n}$ in this setting is the years of education, which, according to GLLS, is a possible control variable that may cause endogeneity. To account for this, they suggest using the country fixed effects.

We expect that education would be a prospective source of the heterogeneity in spatial correlations of the regional productivity since studies like Moretti (2004), and Dong (2010), investigate and demonstrate the spillover effect of human capital accumulation on total factor productivity. The excluded instruments we adopt in the control function are some of the characteristics in GLLS data but are not included in the main regression. Excluded instruments are required for identification because the years of education also enters the EHSAR model as a regressor [13]. Those variables include the probability that residents inside the country speak the same language, the ln(density) of the electric power grid, and the ln(travel times). They are included in the supplementary data set of GLLS, considered

11. These subnational regions are the administrative divisions that can implement their own policies. For instance, they are states in the US and provinces in China.

12. Rook contiguity is obtained using GeoDa by the location data in Sanso-Navarro et al. (2017).

13. Since, otherwise, the control variable $\varepsilon_n = Z_n - X_{2n}\Gamma$ would have perfect multicollinearity with the observables X_{1n}.

exogenous, and not in the main regressions of Sanso-Navarro et al. (2017) and Hoshino (2022). We believe that they are correlated with years of education since, intuitively, a unified language system and a well-built electricity grid facilitates the schooling. Besides, the variable travel times between the cities is also a proxy for the local infrastructure [14].

After cleaning and combining the GLLS data set with location information, we are left with slightly more than 1400 subregions. Table 7 shows the estimation results from different model settings. We use different functions for the nonlinear heterogeneity component $\psi(\cdot)$, including CDF for standard normal, Logistic and Gamma distribution. Robustly, as in GLLS and Hoshino (2022), we have positive coefficients in front of the education-related variables, which are the main interests of growth-related literature, though the magnitudes vary with models. The spatial correlation parameter ρ are significant in all settings, confirming the conclusions in both Sanso-Navarro et al. (2017) and Hoshino (2022). Besides, country FE should always be included for a better fit of the data.

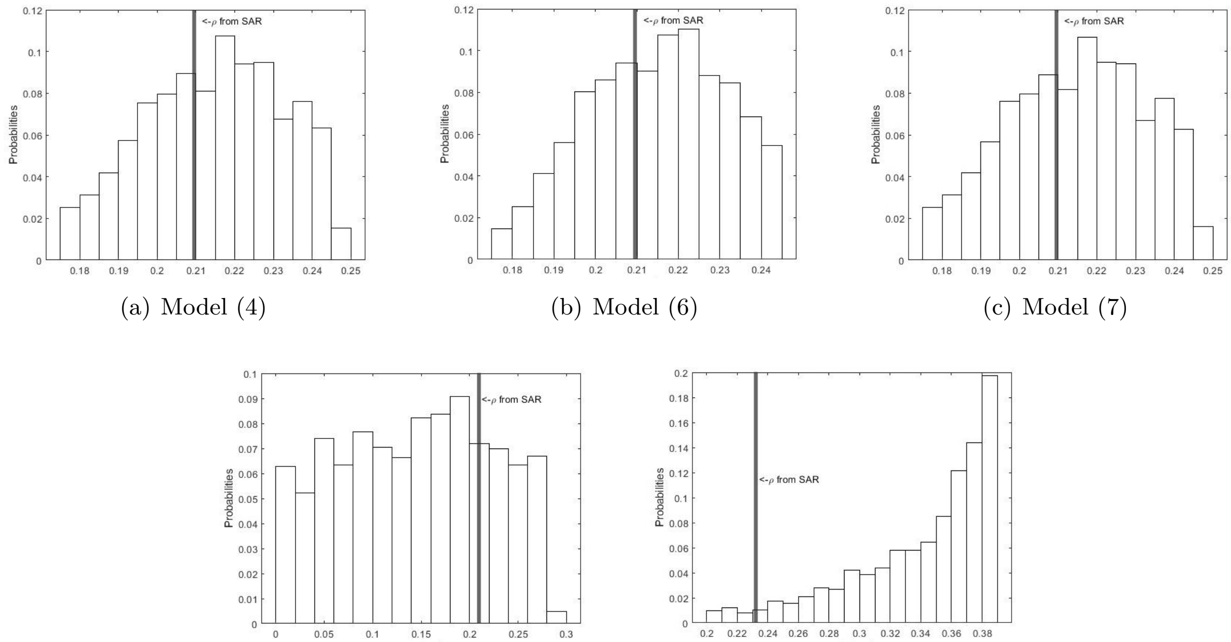
Whether or not we have a significant heterogeneity component depends on the choice of F and the model specification. $\hat{\lambda}$ is positive but insignificant in Model (4), (6), and (7), but a significant heterogeneity component presents if we have F distributed as $Gamma(2,2)$, or, ln(years of education) in place of years of education. Figure 2 contains the distribution of $\hat{\rho}F(z'_{i,n}\hat{\lambda})$ whenever a heterogeneity component is introduced into the SAR model.

Furthermore, for model comparisons, following Jeong et al. (2021), we adopt an Akaike information criterion (AIC) framework. The AIC for a model ℓ equals $-2 \times \log \text{likelihood} + 2M$ (Akaike (1973)), where M is the number of parameters estimated. Jeong et al. (2021) indicates a quantity $\Delta_\ell = \exp((AIC_{true} - AIC_\ell)/2)$, with AIC_{true} as the AIC evaluated at the true model. Jeong et al. (2021) also suggests an Akaike weight $\bar{\Delta}_\ell = \Delta_\ell / \sum_{r=1}^{R} \Delta_r$, where R refers to the candidate models, and the unknown AIC_{true} is cancelled out. $\bar{\Delta}_\ell$ represents the probability that the information loss is minimized at model ℓ, given all other candidates. Table 8 contains the calculated pairwise probabilities led by Akaike weights. For instance, the first cell in Table 8 refers to $\bar{\Delta}_1 = \Delta_1 / \sum_{r=1}^{2} \Delta_r = 0$ and $\bar{\Delta}_2 = \Delta_2 / \sum_{r=1}^{2} \Delta_r = 1$, which means the model with a spatial autoregressive term better captures the underlying pattern of the data [15]. Table 8 shows that $N(0,1)$ and Logistic F functions give almost identical results, whereas they are superior to Model (8). Moreover, Akaike weights favors Model (9)

14. Moreover, running OLS with years of education as dependent variable on all the exogenous variables and instruments give the estimated coefficients for the instruments as – 0.5652(0.1484) for unified language, 0.1816(0.0378) for electricity grid, and $-0.3468(0.0383)$ for the travel time across cities. They are all significant and have the expected signs.

15. We divide the models into two groups in Table 8, since including the control function for a potential endogenous variable years of education make it a different model from the exogenous cases. Thus, Model (1) to Model (3) forms one group, while Model (4), (6), (8), and (9) constitute the other. Model (5) is omitted due to the poor performance in the control function, and Model (7) is almost the same as Model (4).

over Model (4) (6) and (8), where a logged years of education enters in place of the level one.



(a) Model (4) (b) Model (6) (c) Model (7)

(d) Model (8) (e) Model (9)

Figure 2: Distribution of the estimation heterogeneity component, $\hat{\rho}F(z_{i,n}'\hat{\lambda})$, for models in Table 7.

	No Spatial	SAR		N(0,1)		Logistic	N(0,0.5)	Gamma (2,2)	N(0,1)
Models	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	-	-	0.7557***	-	0.6290	-	-	-	-
	-	-	(0.1179)	-	(0.5716)	-	-	-	-
Temperature	-0.0143***	-0.0115***	0.0012	-0.0093	0.0090	-0.0093	-0.0093	-0.0101	-0.0064
	(0.0027)	(0.0026)	(0.0017)	(0.0065)	(0.0077)	(0.0065)	(0.0065)	(0.0065)	(0.0023)
Inverse distance to coast	0.4324***	0.3491***	0.3749***	0.2700	0.2505	0.2700	0.2701	0.3110	0.1097
	(0.0952)	(0.0934)	(0.0749)	(0.2336)	(0.3887)	(0.2336)	(0.2336)	(0.2336)	(0.0872)
ln Oil production per capita	0.2066***	0.1912***	0.1715***	0.1848***	0.1584	0.1848***	0.1849***	0.1929***	0.1729***
	(0.0218)	(0.0213)	(0.0244)	(0.0535)	(0.1302)	(0.0535)	(0.0535)	(0.0533)	(0.0167)
ln Population	0.0150	0.0145	0.0245***	-0.0015	0.0158	-0.0014	-0.0015	-0.0045	0.0135
	(0.0096)	(0.0093)	(0.0062)	(0.0237)	(0.0331)	(0.0237)	(0.0237)	(0.0237)	(0.0090)
Years of education	0.2783***	0.2689***	0.0798***	0.2917***	0.1101***	0.2949***	0.2916***	0.1447***	-
	(0.0101)	(0.0100)	(0.0052)	(0.0389)	(0.0158)	(0.0384)	(0.0388)	(0.0450)	-
ln Years of education	-	-	-	-	-	-	-	-	1.3891***
	-	-	-	-	-	-	-	-	(0.1855)
ln No. ethnic groups	-0.0385**	-0.0327**	-0.0511***	-0.0059	-0.0170	-0.0059	-0.0060	-0.0089	0.0158
	(0.0150)	(0.0146)	(0.0153)	(0.0374)	(0.0808)	(0.0374)	(0.0374)	(0.0374)	(0.0192)
ρ	-	0.2096***	0.7773***	0.3478***	0.7759***	0.3532***	0.3478***	0.4605***	0.3918***
	-	(0.0275)	(0.0100)	(0.0609)	(0.0101)	(0.0605)	(0.0609)	(0.0772)	(0.0512)
λ	-	-	-	0.0431	12.0691	0.0630	0.0305	0.3240***	0.1864***
	-	-	-	(0.0270)	(16.6219)	(0.0404)	(0.0191)	(0.0415)	(0.0553)
δ	-	-	-	-0.0793**	-0.0363***	-0.0792**	-0.0793**	-0.0852***	-0.9277***
	-	-	-	(0.0315)	(0.0121)	(0.0315)	(0.0315)	(0.0318)	(0.1826)
Country FE	Y	Y	N	Y	N	Y	Y	Y	Y
Sample Size	1453	1453	1453	1433	1433	1433	1433	1433	1433
# of parameters Dummies Included	112	113	9	226	20	226	226	226	226
Log-Likelihood	1103.7	1131.16	583.817	922.952	-1079.32	922.865	922.952	920.068	2619.349

Standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Estimation results for the GLLS application.

Models	Model (2)	Model (3)	Model (6)	Model (8)	Model (9)
Model (1)	(0.0000,1.0000)	(1.0000,0.0000)	-	-	-
Model (2)	-	(1.0000,0.0000)	-	-	-
Model (4)	-	-	(0.5217, 0.4783)	(0.9470, 0.0530)	(0.0000,1.0000)
Model (6)	-	-	-	(0.9425, 0.0575)	(0.0000,1.0000)
Model (8)	-	-	-	-	(0.0000,1.0000)

Table 8: The pairwise Akaike weights for Models in Table 7.

6 Conclusion

In this paper, we generalize the traditional cross-sectional SAR model with a linear interaction term to an EHSAR model which accounts for nonlinear endogenous heterogeneity in spatial spillovers. We provide the specifications for the EHSAR outcome equation and the equation for endogenous entries in the heterogeneity component, and propose a QML estimation method combined with a control function approach. The asymptotic properties of the QMLE are established using the theory of asymptotic inference under near-epoch dependence, and the theoretical results in finite samples are verified by a Monte Carlo simulation study. By applying our model to the regional productivity in the GLLS data, we identify years of education as an endogenous heterogeneity source in income spillovers.

Appendix

Appendix A. QMLE derivation

The first order derivatives are

$$
\begin{aligned}
\frac{\partial \ln L_n(\theta)}{\partial \lambda_p} &= \frac{1}{\sigma_\xi^2}\left[\operatorname{diag}\left(\frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_p}\right) W_n Y_n\right]' \xi_n(\theta) - \operatorname{tr}\left[\operatorname{diag}\left(\frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_p}\right) W_n S_n^{-1}(\Lambda)\right], p = 1, ..., P;\\
\frac{\partial \ln L_n(\theta)}{\partial \beta} &= \frac{1}{\sigma_\xi^2} X_{1n}' \xi_n(\theta);\\
\frac{\partial \ln L_n(\theta)}{\partial \operatorname{vec}(\Gamma)} &= \left(\varSigma_\varepsilon^{-1} \otimes X_{2n}'\right) \operatorname{vec}(Z_n - X_{2n}\Gamma) - \frac{1}{\sigma_\xi^2} \delta \otimes \left(X_{2n}' \xi_n(\theta)\right);\\
\frac{\partial \ln L_n(\theta)}{\partial \sigma_\xi^2} &= -\frac{n}{2\sigma_\xi^2} + \frac{1}{2\sigma_\xi^4} \xi_n(\theta)' \xi_n(\theta);\\
\frac{\partial \ln L_n(\theta)}{\partial \alpha} &= -\frac{n}{2}\frac{\partial \ln|\varSigma_\varepsilon|}{\partial \alpha} - \frac{1}{2}\frac{\partial}{\partial \alpha} \operatorname{tr}\left[\varSigma_\varepsilon^{-1} \varepsilon_n(\Gamma)' \varepsilon_n(\Gamma)\right];\\
\frac{\partial \ln L_n(\theta)}{\partial \delta} &= \frac{1}{\sigma_\xi^2} \varepsilon_n(\theta)' \xi_n(\theta),
\end{aligned}
$$

where $\xi_n(\theta) = S_n(\Lambda) Y_n - X_{1n}\beta - (Z_n - X_{2n}\Gamma)\delta$ and $\varepsilon_n(\Gamma) = Z_n - X_{2n}\Gamma$. As α is a J-dimensional column vector of distinct elements in Σ_ε, the J-dimensional vector $\frac{\partial \ln|\Sigma_\varepsilon|}{\partial\alpha}$ has the jth element $\mathrm{tr}\left(\Sigma_\varepsilon^{-1}\frac{\partial\Sigma_\varepsilon}{\partial\alpha_j}\right)$ and $\frac{\partial}{\partial\alpha}\mathrm{tr}\left[\Sigma_\varepsilon^{-1}\varepsilon_n(\Gamma)'\varepsilon_n(\Gamma)\right]$ has its jth element $-\mathrm{tr}\left(\Sigma_\varepsilon^{-1}\frac{\partial\Sigma_\varepsilon}{\partial\alpha_j}\Sigma_\varepsilon^{-1}\varepsilon_n(\Gamma)'\varepsilon_n(\Gamma)\right)$ for $j = 1, ..., J$.

The second order derivatives are

$$
\begin{aligned}
\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda_p\partial\lambda_p} &= -\frac{1}{\sigma_\xi^2}\left[\mathrm{diag}\left(\frac{\partial\psi(\Lambda, z_{i,n})}{\partial\lambda_p}\right)W_nY_n\right]'\left[\mathrm{diag}\left(\frac{\partial\psi(\Lambda, z_{i,n})}{\partial\lambda_p}\right)W_nY_n\right] \\
&\quad - \mathrm{tr}\left[\mathrm{diag}\left(\frac{\partial^2\psi(\Lambda, z_{i,n})}{\partial\lambda_p\partial\lambda_p}\right)W_nS_n^{-1}(\Lambda)\right] - \mathrm{tr}\left[\mathrm{diag}\left(\frac{\partial\psi(\Lambda, z_{i,n})}{\partial\lambda_p}\right)W_nS_n^{-1}(\Lambda)\right]^2, \ p = 1, ..., P; \\
\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda_p\partial\lambda_q} &= -\frac{1}{\sigma_\xi^2}\left[\mathrm{diag}\left(\frac{\partial\psi(\Lambda, z_{i,n})}{\partial\lambda_p}\right)W_nY_n\right]'\left[\mathrm{diag}\left(\frac{\partial\psi(\Lambda, z_{i,n})}{\partial\lambda_q}\right)W_nY_n\right] - \mathrm{tr}\left[\mathrm{diag}\left(\frac{\partial^2\psi(\Lambda, z_{i,n})}{\partial\lambda_p\partial\lambda_q}\right)W_nS_n^{-1}(\Lambda)\right] \\
&\quad - \mathrm{tr}\left[\mathrm{diag}\left(\frac{\partial\psi(\Lambda, z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda, z_{i,n})}{\partial\lambda_q}\right)(W_nS_n^{-1}(\Lambda))^2\right], \ p \neq q \text{ and } p, q = 1, ..., P; \\
\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda_p\partial\beta} &= -\frac{1}{\sigma_\xi^2}X_{1n}'\left[\mathrm{diag}\left(\frac{\partial\psi(\Lambda, z_{i,n})}{\partial\lambda_p}\right)W_nY_n\right]; \ \frac{\partial^2 \ln L_n(\theta)}{\partial\lambda_p\partial\mathrm{vec}(\Gamma)} = \frac{1}{\sigma_\xi^2}\delta\otimes\left(X_{2n}'\left[\mathrm{diag}\left(\frac{\partial\psi(\Lambda, z_{i,n})}{\partial\lambda_p}\right)W_nY_n\right]\right); \\
\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda_p\partial\sigma_\xi^2} &= -\frac{1}{\sigma_\xi^4}\left[\mathrm{diag}\left(\frac{\partial\psi(\Lambda, z_{i,n})}{\partial\lambda_p}\right)W_nY_n\right]'\xi_n(\theta); \ \frac{\partial^2 \ln L_n(\theta)}{\partial\lambda_p\partial\alpha} = 0; \\
\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda_p\partial\delta} &= -\frac{1}{\sigma_\xi^2}\varepsilon_n(\Gamma)'\left[\mathrm{diag}\left(\frac{\partial\psi(\Lambda, z_{i,n})}{\partial\lambda_p}\right)W_nY_n\right]; \\
\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\beta'} &= -\frac{1}{\sigma_\xi^2}X_{1n}'X_{1n}; \ \frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\mathrm{vec}(\Gamma)'} = \frac{1}{\sigma_\xi^2}\delta\otimes(X_{2n}'X_{1n}); \ \frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\sigma_\xi^2} = -\frac{1}{\sigma_\xi^4}X_{1n}'\xi_n(\theta); \\
\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\alpha'} &= 0; \ \frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\delta'} = -\frac{1}{\sigma_\xi^2}X_{1n}'\varepsilon_n(\Gamma); \\
\frac{\partial^2 \ln L_n(\theta)}{\partial\mathrm{vec}(\Gamma)\partial\mathrm{vec}(\Gamma)'} &= -\Sigma_\varepsilon^{-1}\otimes(X_{2n}'X_{2n}) - \frac{1}{\sigma_\xi^2}\delta\delta'\otimes(X_{2n}'X_{2n}); \ \frac{\partial^2 \ln L_n(\theta)}{\partial\mathrm{vec}(\Gamma)\partial\sigma_\xi^2} = \frac{1}{\sigma_\xi^4}\delta\otimes(X_{2n}'\xi_n(\theta)); \\
\frac{\partial^2 \ln L_n(\theta)}{\partial\mathrm{vec}(\Gamma)\partial\alpha'} &= \left[I_h\otimes(X_{2n}'\varepsilon_n(\Gamma))\right]\frac{\partial\mathrm{vec}(\Sigma_\varepsilon^{-1})}{\partial\alpha'}; \ \frac{\partial^2 \ln L_n(\theta)}{\partial\mathrm{vec}(\Gamma)\partial\delta'} = -\frac{1}{\sigma_\xi^2}I_h\otimes(X_{2n}'\xi_n(\theta)) + \frac{1}{\sigma_\xi^2}\delta\otimes(X_{2n}'\varepsilon_n(\Gamma)); \\
\frac{\partial^2 \ln L_n(\theta)}{\partial\sigma_\xi^2\partial\sigma_\xi^2} &= \frac{n}{2\sigma_\xi^4} - \frac{1}{\sigma_\xi^6}\xi_n(\theta)'\xi_n(\theta); \ \frac{\partial^2 \ln L_n(\theta)}{\partial\sigma_\xi^2\partial\alpha} = 0; \ \frac{\partial^2 \ln L_n(\theta)}{\partial\sigma_\xi^2\partial\delta} = -\frac{1}{\sigma_\xi^4}\varepsilon_n(\theta)'\xi_n(\theta); \\
\frac{\partial^2 \ln L_n(\theta)}{\partial\alpha\partial\alpha'} &= -\frac{n}{2}\frac{\partial^2\ln|\Sigma_\varepsilon|}{\partial\alpha\partial\alpha'} - \frac{1}{2}\frac{\partial^2\mathrm{tr}[\Sigma_\varepsilon^{-1}\varepsilon_n(\Gamma)'\varepsilon_n(\Gamma)]}{\partial\alpha\partial\alpha'}; \ \frac{\partial^2 \ln L_n(\theta)}{\partial\alpha\partial\delta'} = 0; \ \frac{\partial^2 \ln L_n(\theta)}{\partial\delta\partial\delta'} = -\frac{1}{\sigma_\xi^2}\varepsilon_n(\theta)'\varepsilon_n(\theta).
\end{aligned}
$$

where $\frac{\partial^2\ln|\Sigma_\varepsilon|}{\partial\alpha\partial\alpha'}$ is a $J\times J$ matrix with the (j,k)th element $\frac{\partial^2\ln|\Sigma_\varepsilon|}{\partial\alpha_j\partial\alpha_k} = -\mathrm{tr}\left(\Sigma_\varepsilon^{-1}\frac{\partial\Sigma_\varepsilon}{\partial\alpha_k}\Sigma_\varepsilon^{-1}\frac{\partial\Sigma_\varepsilon}{\partial\alpha_j}\right)$ and the (j,k)th element of $\frac{\partial^2\mathrm{tr}[\Sigma_\varepsilon^{-1}\varepsilon_n(\Gamma)'\varepsilon_n(\Gamma)]}{\partial\alpha\partial\alpha'}$ is $\frac{\partial^2}{\partial\alpha_j\partial\alpha_k}\mathrm{tr}[\Sigma_\varepsilon^{-1}\varepsilon_n(\Gamma)'\varepsilon_n(\Gamma)] = \mathrm{tr}(\Sigma_\varepsilon^{-1}(\frac{\partial\Sigma_\varepsilon}{\partial\alpha_k}\Sigma_\varepsilon^{-1}\frac{\partial\Sigma_\varepsilon}{\partial\alpha_j} + \frac{\partial\Sigma_\varepsilon}{\partial\alpha_j}\Sigma_\varepsilon^{-1}\frac{\partial\Sigma_\varepsilon}{\partial\alpha_k})\times$ $\Sigma_\varepsilon^{-1}\varepsilon_n(\Gamma)'\varepsilon_n(\Gamma))$.

Therefore,

$$
E\left(\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}\right)
$$

$$
= -\frac{1}{\sigma^2_{\xi,0}}\begin{bmatrix}
a_{11} & a_{12} & \dots & a_{1P} & E(\widetilde{G}_{n,1}\tilde{X}_n)'X_{1n} & A_{\lambda_1\Gamma} & E[\mathrm{tr}(\widetilde{G}_{n,1})] & 0 & E(\widetilde{G}_{n,1}\tilde{X}_n)'\varepsilon_n \\
* & a_{22} & \dots & a_{2P} & E(\widetilde{G}_{n,2}\tilde{X}_n)'X_{1n} & A_{\lambda_2\Gamma} & E[\mathrm{tr}(\widetilde{G}_{n,2})] & 0 & E(\widetilde{G}_{n,2}\tilde{X}_n)'\varepsilon_n \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
* & * & \dots & a_{PP} & E(\widetilde{G}_{n,P}\tilde{X}_n)'X_{1n} & A_{\lambda_P\Gamma} & E[\mathrm{tr}(\widetilde{G}_{n,P})] & 0 & E(\widetilde{G}_{n,P}\tilde{X}_n)'\varepsilon_n \\
* & * & \dots & * & X'_{1n}X_{1n} & -\delta'_0 \otimes (X'_{1n}X_{2n}) & 0 & 0 & 0 \\
* & * & \dots & * & * & A_{\Gamma\Gamma} & 0 & 0 & 0 \\
* & * & \dots & * & * & * & \frac{n}{2\sigma^2_{\xi,0}} & 0 & 0 \\
* & * & \dots & * & * & * & * & A_{\alpha\alpha} & 0 \\
* & * & \dots & * & * & * & * & * & n\Sigma_{\varepsilon,0}
\end{bmatrix}
$$

where $\tilde{X}_n = X_{1n}\beta_0 + \varepsilon_n\delta_0$, $G_n = W_nS_n^{-1}$, $\widetilde{G}_{n,p} = \mathrm{diag}(\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p})G_n$, $\widetilde{G}_{n,pp} = \mathrm{diag}(\frac{\partial^2\psi(\Lambda,z_{i,n})}{\partial\lambda_p\partial\lambda_p})G_n$, $\widetilde{G}_{n,pq} = \mathrm{diag}(\frac{\partial^2\psi(\Lambda,z_{i,n})}{\partial\lambda_p\partial\lambda_q})G_n$, $p = 1,..,P$; $a_{pp} = E[(\widetilde{G}_{n,p}\tilde{X}_n)'(\widetilde{G}_{n,p}\tilde{X}_n)] + \sigma^2_{\xi,0}E[\mathrm{tr}(\widetilde{G}'_{n,p}\widetilde{G}_{n,p}) + \mathrm{tr}(\widetilde{G}^2_{n,p} + \widetilde{G}_{n,pp})]$, $p = 1,..,P$; $a_{pq} = E(\widetilde{G}_{n,p}\tilde{X}_n)'(\widetilde{G}_{n,q}\tilde{X}_n) + \sigma^2_{\xi,0}E[\mathrm{tr}(\widetilde{G}'_{n,p}\widetilde{G}_{n,q}) + \mathrm{tr}(\widetilde{G}'_{n,p}\widetilde{G}_{n,q} + \widetilde{G}_{n,pq})]$, $p \neq q, p, q = 1,..,P$; $A_{\lambda_p\Gamma} = -\delta'_0 \otimes \left[E\left(\widetilde{G}_{n,p}\tilde{X}_n\right)' X_{2n}\right], p = 1,...,P$; $A_{\Gamma\Gamma} = \left(\sigma^2_{\xi,0}\Sigma^{-1}_{\varepsilon,0} + \delta_0\delta'_0\right) \otimes (X'_{2n}X_{2n})$; $(A_{\alpha\alpha})_{kj} = \frac{n\sigma^2_{\xi 0}}{2}\,\mathrm{tr}\left(\Sigma^{-1}_{\varepsilon 0}\frac{\partial\Sigma_{\varepsilon 0}}{\partial\alpha_k}\Sigma^{-1}_{\varepsilon 0}\frac{\partial\Sigma_{\varepsilon 0}}{\partial\alpha_j}\right)$ for $j, k = 1,...,J$. And $E\left(\frac{\partial \ln L_n(\theta_0)}{\partial\theta}\frac{\partial \ln L_n(\theta_0)}{\partial\theta'}\right) = -E\left(\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}\right) + \Omega^{ML}_{\theta_0}$ with

$$
\Omega^{ML}_{\theta_0}
$$

$$
= \frac{1}{\sigma^4_{\xi,0}} \cdot \begin{bmatrix}
b_{11} & b_{12} & \dots & b_{1P} & (B_{\lambda_p\beta'})_1 & -E[\mu_{3,0}\delta'_0 \otimes \mathrm{vec}'_D(\widetilde{G}_{n,1})X_{2n}] & (B_{\lambda_p\sigma^2_\xi})_1 & 0 & (B_{\lambda_p\delta'})_1 \\
* & b_{22} & \dots & b_{2P} & (B_{\lambda_p\beta'})_2 & -E[\mu_{3,0}\delta'_0 \otimes \mathrm{vec}'_D(\widetilde{G}_{n,2})X_{2n}] & (B_{\lambda_p\sigma^2_\xi})_2 & 0 & (B_{\lambda_p\delta'})_2 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
* & * & \dots & b_{PP} & (B_{\lambda_p\beta'})_P & -E[\mu_{3,0}\delta'_0 \otimes \mathrm{vec}'_D(\widetilde{G}_{n,2})X_{2n}] & (B_{\lambda_p\sigma^2_\xi})_P & 0 & (B_{\lambda_p\delta'})_P \\
* & * & \dots & * & 0 & 0 & \frac{\mu_{3,0}}{2\sigma^2_{\xi,0}}X'_{1n}l_n & 0 & 0 \\
* & * & \dots & * & * & 0 & -\frac{\mu_{3,0}}{2\sigma^2_{\xi,0}}\delta_0 \otimes (X'_{2n}l_n) & B_{\Gamma\alpha} & 0 \\
* & * & \dots & * & * & * & \frac{n(\mu_{4,0}-3\sigma^4_{\xi,0})}{4\sigma^4_{\xi,0}} & 0 & \frac{\mu_{3,0}}{2\sigma^2_{\xi,0}}\varepsilon'_n l_n \\
* & * & \dots & * & * & * & * & B_{\alpha\alpha} & 0 \\
* & * & \dots & * & * & * & * & * & 0
\end{bmatrix}
$$

where $b_{pp} = E[2\mu_{3,0}(\widetilde{G}_{n,p}\tilde{X}_n)'\mathrm{vec}_D(\widetilde{G}_{n,p}) + (\mu_{4,0} - 3\sigma_{\xi,0}^4)\mathrm{vec}_D'(\widetilde{G}_{n,p})\mathrm{vec}_D(\widetilde{G}_{n,p})]$, $p = 1, ..., P$; $b_{pq} = E[\mu_{3,0}(\widetilde{G}_{n,p}\tilde{X}_n)'\mathrm{vec}_D(\widetilde{G}_{n,q}) + \mu_{3,0}(\widetilde{G}_{n,q}\tilde{X}_n)'\mathrm{vec}_D(\widetilde{G}_{n,p}) + (\mu_{4,0} - 3\sigma_{\xi,0}^4)\mathrm{vec}_D'(\widetilde{G}_{n,p})\mathrm{vec}_D(\widetilde{G}_{n,q})]$, $p \neq q, p, q = 1, ..., P$; $(B_{\lambda_p\beta'})_p = E[\mu_{3,0}\mathrm{vec}_D'(\widetilde{G}_{n,p})X_{1n}]$, $(B_{\lambda_p\sigma_\xi^2})_p = \frac{1}{2\sigma_{\xi,0}^2}E[\mu_{3,0}(\widetilde{G}_{n,1}\tilde{X}_n)'l_n + (\mu_{4,0} - 3\sigma_{\xi,0}^4)\mathrm{tr}(\widetilde{G}_{n,1})]$, $(B_{\lambda_p\delta'})_p = E[\mu_{3,0}\mathrm{vec}_D'(\widetilde{G}_{n,p})\varepsilon_n]$, $p = 1, ..., P$; $B_{\Gamma\alpha} = \frac{\sigma_{\xi,0}^4}{2}[l_n' \otimes E(\varepsilon_{i,n}'\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_j}\Sigma_{\varepsilon 0}^{-1}\varepsilon_{i,n} \times \Sigma_{\varepsilon 0}^{-1}\varepsilon_{i,n}) \otimes I_{k_2}]\,\mathrm{vec}(X_{2n}')$; $(B_{\alpha\alpha})_{kj} = \frac{n\sigma_{\xi 0}^4}{4}[E(\varepsilon_{i,n}'\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_j}\Sigma_{\varepsilon 0}^{-1}\varepsilon_{i,n}\varepsilon_{i,n}'\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_k}\Sigma_{\varepsilon 0}^{-1}\varepsilon_{i,n}) - \mathrm{tr}\left(\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_j}\right)\mathrm{tr}(\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_k}) - 2\,\mathrm{tr}(\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_k}\Sigma_{\varepsilon 0}^{-1}\frac{\partial \Sigma_{\varepsilon 0}}{\partial \alpha_j})]$, $k, j = 1, .., J$.

Appendix B. Mathematical proofs

Proof of Proposition 1. Denote $|A| = (|a_{ij}|)$ for any matrix $A = (a_{ij})$.

(i) As $Y_n = [I_n - \Psi(\Lambda_0, Z_n)W_n]^{-1}(X_{1n}\beta_0 + (Z_n - X_{2n}\Gamma_0)\delta_0 + \xi_n)$ and $[I_n - \Psi(\Lambda_0, Z_n)W_n]^{-1} = \sum_{l=0}^{\infty}[\Psi(\Lambda_0, Z_n)W_n]^l \leq^* \sum_{l=0}^{\infty}|b_\psi W_n|^l = (I_n - |b_\psi W_n|)^{-1} \equiv M_n \equiv (m_{ij,n})$, where $A \leq^* B$ means $|a_{ij}| \leq |b_{ij}|$ for all i's and j's, we have $|y_{i,n}| \leq \sum_{j=1}^n m_{ij,n}\,|x_{1,jn}\beta_0 + \varepsilon_{j,n}\delta_0 + \xi_{j,n}|$. Then, by the Minkowski's inequality, $\|y_{i,n}\|_\eta \leq \sum_{j=1}^n m_{ij,n}\|x_{1,jn}\beta_0 + \varepsilon_{j,n}\delta_0 + \xi_{j,n}\|_\eta$. With the uniform L_η-boundedness of $\{y_{i,n}\}_{i=1}^n$, the uniform L_η-boundedness of $\{\psi(\Lambda_0, z_{i,n})w_{i\cdot,n}Y_n\}_{i=1}^n$ can be obtained by applying the Minkowski's inequality.

(ii) First, we consider the NED property of $\{y_{i,n}\}_{i=1}^n$. Let $Y_n^{(1)} = \Psi(\Lambda_0, Z_n^{(1)})W_nY_n^{(1)} + X_{1n}^{(1)}\beta_0 + (Z_n^{(1)} - X_{2n}^{(1)}\Gamma_0)\delta_0 + \xi_n^{(1)}$ and $Y_n^{(2)} = \Psi(\Lambda_0, Z_n^{(2)})W_nY_n^{(2)} + X_{1n}^{(2)}\beta_0 + (Z_n^{(2)} - X_{2n}^{(2)}\Gamma_0)\delta_0 + \xi_n^{(2)}$, from the proof of part (i), we have $|y_{i,n}^{(1)} - y_{i,n}^{(2)}| \leq \sum_{j=1}^n m_{ij,n}|(x_{1,jn}^{(1)} - x_{1,jn}^{(2)})'\beta_0 + (\epsilon_{j,n}^{(1)} - \epsilon_{j,n}^{(2)})'\delta_0 + (\xi_{j,n}^{(1)} - \xi_{j,n}^{(2)})|$, where $(m_{ij,n}) \equiv (I_n - |b_\psi W_n|)^{-1}$. Then by Proposition 1 and its proof in Jenish and Prucha (2012), we have $\|y_{i,n} - E[y_{i,n}|\mathcal{F}_{i,n}(s)]\|_2 \leq \sup_{j,n}\|\xi_{j,n} + x_{1,jn}'\beta_0 + \varepsilon_{j,n}'\delta_0\|_2 \sup_{i,n}\sum_{j:\rho_{ij}>s} m_{ij,n}$. Under Assumption 4, by Claim C.1.2, Claim C.1.6 and its proof in Qu and Lee (2015), we have

$$\sum_{k_1}\cdots\sum_{k_{l-1}}|w_{ik_1,n}w_{k_1k_2,n}\cdots w_{k_{l-2}k_{l-1}}w_{k_{l-1}j,n}| \leq c_3 l^{c_2d_0+2}c_w^l\rho_{ij}^{-c_2d_0}$$

for some constant $c_3 > 0$. Then, for any $j \neq i$,

$$\begin{aligned}|(I_n - |b_\psi W_n|)_{ij}^{-1}| &= \sum_{l=1}^{\infty}|b_\psi W_n|_{ij}^l = \sum_{l=1}^{\infty}|b_\psi^l|\sum_{k_1}\cdots\sum_{k_{l-1}}|w_{ik_1,n}w_{k_1k_2,n}\cdots w_{k_{l-2}k_{l-1}}w_{k_{l-1}j,n}| \\ &\leq c_3\rho_{ij}^{-c_2d_0}\sum_{l=1}^{\infty}|b_\psi c_w|^l l^{c_2d_0+2} \leq c_4\rho_{ij}^{-c_2d_0}\end{aligned}$$

for some constant $c_4 > 0$. As $|\{j : m \leq \rho_{ij} < m+1\}| \leq c_5 m^{d_0-1}$ for some constant $c_5 > 0$ by Lemma

A.1 in Jenish and Prucha (2009), then when s is large enough,

$$\begin{aligned}
&\sup_{i,n}\sum_{j:\rho_{ij}>s} m_{ij,n} \le \sup_{i,n}\sum_{m=[s]}^{\infty}\sum_{j:m\le\rho_{ij}<m+1} c_4\rho_{ij}^{-c_2d_0} \le \sum_{m=[s]}^{\infty} c_5 m^{d_0-1}c_4 m^{-c_2d_0}\\
&\quad\le \sum_{m=[s]}^{\infty} c_5c_4(m+1)^{d_0-1}[(m+1)/2]^{-c_2d_0} \le c_5c_4 2^{-c_2d_0}\int_s^{\infty} x^{-c_2d_0+d_0-1}dx = c_5c_4 2^{c_2d_0}(c_2d_0-d_0)^{-1}s^{(1-c_2)d_0}
\end{aligned}$$

which implies that $\|y_{i,n}-E[y_{i,n}|\mathcal{F}_{i,n}(s)]\|_2 \le c_4c_5 2^{c_2d_0}(c_2d_0-d_0)^{-1}s^{(1-c_2)d_0}\sup_{j,n}\|\xi_{j,n}+x_{1,jn}'\beta_0+\varepsilon_{j,n}'\delta_0\|_2 \le Cs^{(1-c_2)d_0}$ for some constant $C>0$.

Next, we consider the NED of $\{\psi(\Lambda_0,z_{i,n})w_{i\cdot,n}Y_n\}_{i=1}^n$. Because $|\{j : m\le\rho_{ij}<m+1\}|\le c_5m^{d_0-1}$ for some constant $c_5>0$, when s is large enough,

$$\begin{aligned}
&\|\psi(\Lambda_0,z_{i,n})w_{i\cdot,n}Y_n - E[\psi(\Lambda_0,z_{i,n})w_{i\cdot,n}Y_n|\mathcal{F}_{i,n}(s)]\|_2\\
&\quad\le \sum_{k:\rho_{ki}\le s/2}|b_\psi||w_{ik,n}|\cdot\|y_{k,n}-E[y_{k,n}|\mathcal{F}_{i,n}(s)]\|_2 + \sum_{k:\rho_{ki}>s/2}|b_\psi||w_{ik,n}|\cdot\|y_{k,n}-E[y_{k,n}|\mathcal{F}_{i,n}(s)]\|_2\\
&\quad\le \sum_{k:\rho_{ki}\le s/2}|b_\psi||w_{ik,n}|\cdot\|y_{k,n}-E[y_{k,n}|\mathcal{F}_{i,n}(s)]\|_2 + \sum_{m=[s/2]}^{\infty}\sum_{k:m\le\rho_{ki}<m+1}|b_\psi||w_{ik,n}|\cdot\|y_{k,n}\|_2\\
&\quad\le \sum_{k:\rho_{ki}\le s/2}|b_\psi||w_{ik,n}|C(s/2)^{(1-c_2)d_0} + (\sup_{n,k}\|y_{k,n}\|_2)\sum_{m=[s/2]}^{\infty} c_5m^{d_0-1}c_4m^{-c_2d_0}\\
&\quad\le b_\psi c_w C(s/2)^{(1-c_2)d_0} + (\sup_{n,k}\|y_{k,n}\|_2)c_5c_4 2^{c_2d_0-d_0+1}\int_{[s/2]}^{\infty}\frac{dx}{x^{c_2d_0-d_0+1}} \le c_6 s^{(1-c_2)d_0}
\end{aligned}$$

for some constant $c_6>0$, where the second inequality comes from $\mathcal{F}_{k,n}(s/2)\subseteq\mathcal{F}_{i,n}(s)$ when $\rho_{ki}\le s/2$.

Proof of Lemma 1. Let $\theta_0 = \left(\Lambda_0', \beta_0', \text{vec}\left(\Gamma_0\right)', \sigma_{\xi,0}^2, \alpha_0', \delta_0'\right)'$ be the true parameter, and $\theta = \left(\Lambda', \beta', \text{vec}\left(\Gamma\right)', \sigma_{\xi}^2, \alpha', \delta'\right)'$ be an arbitrary value of parameter in Θ defined in Assumption 3. Since $\ln x \le x-1$ implies $\ln x \le 2\left(\sqrt{x}-1\right)$ for any $x>0$, we have

$$\begin{aligned}
E\ln\left[L_n(\theta)/L_n\left(\theta_0\right)\right] &\le 2E\left(\sqrt{L_n(\theta)/L_n\left(\theta_0\right)}-1\right) = 2\int\left(\sqrt{L_n(\theta)/L_n\left(\theta_0\right)}-1\right)L_n\left(\theta_0\right)du_n\\
&= 2\int\sqrt{L_n(\theta)L_n\left(\theta_0\right)}du_n - 1 = -\int\left[\sqrt{L_n(\theta)}-\sqrt{L_n\left(\theta_0\right)}\right]^2 du_n \le 0
\end{aligned}$$

where $u_n = \left(Y_n', \text{vec}\left(Z_n\right)'\right)'$.This implies in particular the information inequality $E\ln L_n\left(\theta\right)\le E\ln L_n\left(\theta_0\right)$ for all θ. Thus, θ_0 is a maximizer. Also, this inequality also implies that if $E\ln L_n\left(\theta\right) = E\ln L_n\left(\theta_0\right)$, $\ln L_n\left(\theta\right) = \ln L_n\left(\theta_0\right)$ almost surely. Assume there exists $\theta_1\neq\theta_0$ such that $\ln L_n\left(\theta_1\right) = \ln L_n\left(\theta_0\right)$, which implies

$$
\begin{aligned}
&-\frac{n}{2}\ln\sigma_{\xi,1}^2+\ln|S_n(\Lambda_1)|-\frac{n}{2}\ln|\mathit{\Sigma}_{\varepsilon,1}|-\frac{1}{2}\sum_{i=1}^{n}\left(z_{i,n}'-x_{2,in}'\mathit{\Gamma}_1\right)\mathit{\Sigma}_{\varepsilon,1}^{-1}\left(z_{i,n}-\mathit{\Gamma}_1'x_{2,in}\right)\\
&-\frac{1}{2\sigma_{\xi,1}^2}\left[S_n\left(\Lambda_1\right)Y_n-X_{1n}\beta_1-\left(Z_n-X_{2n}\Gamma_1\right)\delta_1\right]'\left[S_n\left(\Lambda_1\right)Y_n-X_{1n}\beta_1-\left(Z_n-X_{2n}\Gamma_1\right)\delta_1\right]\\
=&-\frac{n}{2}\ln\sigma_{\xi,0}^2+\ln|S_n(\mathit{\Lambda}_0)|-\frac{n}{2}\ln|\mathit{\Sigma}_{\varepsilon,0}|-\frac{1}{2}\sum_{i=1}^{n}\left(z_{i,n}'-x_{2,in}'\mathit{\Gamma}_0\right)\mathit{\Sigma}_{\varepsilon,0}^{-1}\left(z_{i,n}-\mathit{\Gamma}_0'x_{2,in}\right)\\
&-\frac{1}{2\sigma_{\xi,0}^2}\left[S_nY_n-X_{1n}\beta_0-\left(Z_n-X_{2n}\mathit{\Gamma}_0\right)\delta_0\right]'\left[S_nY_n-X_{1n}\beta_0-\left(Z_n-X_{2n}\mathit{\Gamma}_0\right)\delta_0\right]
\end{aligned}
\tag{A.1}
$$

holds for Y_n and Z_n almost surely. Differentiate (A.1) with respect to Y_n, we get

$$
\frac{1}{\sigma_{\xi,1}^2}S_n\left(\Lambda_1\right)'\left[S_n\left(\Lambda_1\right)Y_n-X_{1n}\beta_1-\left(Z_n-X_{2n}\Gamma_1\right)\delta_1\right]=\frac{1}{\sigma_{\xi,0}^2}S_n\left(\Lambda_0\right)'\left[S_nY_n-X_{1n}\beta_0-\left(Z_n-X_{2n}\Gamma_0\right)\delta_0\right]
\tag{A.2}
$$

Differentiate (A.2) with respect to Y_n once more,

$$
\frac{1}{\sigma_{\xi,1}^2}S_n\left(\Lambda_1\right)'S_n\left(\Lambda_1\right)=\frac{1}{\sigma_{\xi,0}^2}{S_n}'S_n
\tag{A.3}
$$

Since $S_n\left(\Lambda\right)=I_n-\Psi\left(\Lambda,Z_n\right)W_n$, where $\Psi\left(\Lambda,Z_n\right)\equiv\operatorname{diag}\left(\psi\left(\Lambda,z_{i,n}\right)\right)$, due to the stochasticity of Z_n, equation (A.3) implies $\sigma_{\xi,1}^2=\sigma_{\xi,0}^2$. Thus, we should have $S_n\left(\Lambda_1\right)'S_n\left(\Lambda_1\right)={S_n}'S_n$ almost surely, which implies

$$
\begin{aligned}
&\operatorname{diag}\left\{\psi\left(\Lambda_0,z_{i,n}\right)-\psi\left(\Lambda_1,z_{i,n}\right)\right\}W_n+W_n'\operatorname{diag}\left\{\psi\left(\Lambda_0,z_{i,n}\right)-\psi\left(\Lambda_1,z_{i,n}\right)\right\}\\
&-W_n'\operatorname{diag}\left\{\psi^2\left(\Lambda_0,z_{i,n}\right)-\psi^2\left(\Lambda_1,z_{i,n}\right)\right\}W_n=0.
\end{aligned}
$$

As Z_n is stochastic and $W_n\neq0$ while $w_{ii,n}=0$, it must be that $\psi\left(\Lambda_0,z_{i,n}\right)=\psi\left(\Lambda_1,z_{i,n}\right)$ almost surely for any i. By Assumption 3(*ii*), it indicates that $\Lambda_1=\Lambda_0$.

As Y_n and Z_n are stochastic, terms without Y_n and Z_n, with only Z_n, and with both Y_n and Z_n should equal almost surely on LHS and RHS of (A.1), i.e.

$$
\begin{aligned}
&n\ln|\mathit{\Sigma}_{\varepsilon,1}|+\sum_{i=1}^{n}x_{2,in}'\Gamma_1\Sigma_{\varepsilon,1}^{-1}\Gamma_1'x_{2,in}+\frac{1}{\sigma_{\xi,0}^2}\left(X_{2,n}\Gamma_1\delta_1-X_{1n}\beta_1\right)'\left(X_{2n}\Gamma_1\delta_1-X_{1n}\beta_1\right)\\
=&n\ln|\mathit{\Sigma}_{\varepsilon,0}|+\sum_{i=1}^{n}x_{2,in}'\Gamma_0\mathit{\Sigma}_{\varepsilon,0}^{-1}\Gamma_0'x_{2,in}+\frac{1}{\sigma_{\xi,0}^2}\left(X_{2n}\Gamma_0\delta_0-X_{1n}\beta_0\right)'\left(X_{2n}\Gamma_0\delta_0-X_{1n}\beta_1\right)
\end{aligned}
\tag{A.4}
$$

$$\begin{aligned}&\sum_{i=1}^{n}\left(z_{i,n}'\Sigma_{\varepsilon,1}^{-1}z_{i,n}-2z_{i,n}'\Sigma_{\varepsilon,1}^{-1}\Gamma_1'x_{2,in}\right)+\frac{1}{\sigma_{\xi,0}^2}\left[\delta_1'Z_n'Z_n\delta_1+2\delta_1'Z_n'\left(X_{1n}\beta_1-X_{2n}\Gamma_1\delta_1\right)\right]\\=&\sum_{i=1}^{n}\left(z_{i,n}'\Sigma_{\varepsilon,0}^{-1}z_{i,n}-2z_{i,n}'\Sigma_{\varepsilon,0}^{-1}\Gamma_0'x_{2,in}\right)+\frac{1}{\sigma_{\xi,0}^2}\left[\delta_0'Z_n'Z_n\delta_0+2\delta_0'Z_n'\left(X_{1n}\beta_0-X_{2n}\Gamma_0\delta_0\right)\right]\end{aligned}\tag{A.5}$$

$$Y_n'S_n'\left[X_{1n}\beta_1+\left(Z_n-X_{2n}\Gamma_1\right)\delta_1\right]=Y_n'S_n'\left[X_{1n}\beta_0+\left(Z_n-X_{2n}\Gamma_0\right)\delta_0\right]\tag{A.6}$$

for almost every value of Y_n and Z_n given $\sigma_{\xi,1}^2=\sigma_{\xi,0}^2$ and $\Lambda_1=\Lambda_0$. By (A.6), as the probability that $Y_n'S_n'$ equals to zero is less than 1, we must have

$$X_{1n}\beta_1+\left(Z_n-X_{2n}\Gamma_1\right)\delta_1=X_{1,n}\beta_0+\left(Z_n-X_{2n}\Gamma_0\right)\delta_0$$

Differentiate the above equation with respect to Z_n, we can get $\delta_1=\delta_0$ and we also have

$$X_{1n}\beta_1-X_{2n}\Gamma_1\delta_1=X_{1n}\beta_0-X_{2n}\Gamma_0\delta_0\tag{A.7}$$

By (A.7), (A.4) and (A.5) can be simplified

$$n\ln\left|\Sigma_{\varepsilon,1}\right|+\sum_{i=1}^{n}x_{2,in}'\Gamma_1\Sigma_{\varepsilon,1}^{-1}\Gamma_1'x_{2,in}=n\ln\left|\Sigma_{\varepsilon,0}\right|+\sum_{i=1}^{n}x_{2,in}'\Gamma_0\Sigma_{\varepsilon,0}^{-1}\Gamma_0'x_{2,in}\tag{A.8}$$

$$\sum_{i=1}^{n}\left(z_{i,n}'\Sigma_{\varepsilon,1}^{-1}z_{i,n}-2z_{i,n}'\Sigma_{\varepsilon,1}^{-1}\Gamma_1'x_{2,in}\right)=\sum_{i=1}^{n}\left(z_{i,n}'\Sigma_{\varepsilon,0}^{-1}z_{i,n}-2z_{i,n}'\Sigma_{\varepsilon,0}^{-1}\Gamma_0'x_{2,in}\right)\tag{A.9}$$

Differentiate (A.9) twice with respect to $z_{i,n}$, we can easily get $\Sigma_{\varepsilon,1}^{-1}=\Sigma_{\varepsilon,0}^{-1}$. The uniqueness of matrix inverse implies $\Sigma_{\varepsilon,1}=\Sigma_{\varepsilon,0}$. Then we have $\sum_{i=1}^{n}z_{i,n}'\Sigma_{\varepsilon,0}^{-1}\left(\Gamma_1'-\Gamma_0'\right)x_{2,in}=0$ implies $\Gamma_1=\Gamma_0$ due to the stochasticity of Z_n. (A.7) becomes $X_{1n}\left(\beta_1-\beta_0\right)=X_{2n}\Gamma_0\left(\delta_1-\delta_0\right)$. By Assumption 3(*iv*), X_{1n} and X_{2n} are not linear dependent, thus $\beta_1=\beta_0$ and $\delta_1=\delta_0$. As a result, we must have $\theta_1=\theta_0$, i.e., θ_0 is the unique maximizer of $\ln L_n\left(\theta_0\right)$, which can be identified.

Proof of Proposition 2. To show the desired uniform convergence, we need to derive the Taylor expansion[16] of $\ln\left|I_n-\Psi(\Lambda,Z_n)W_n\right|$. Define $g(\Lambda)=\ln\left|I_n-\Psi(\Lambda,Z_n)W_n\right|$ and $\Lambda^0=(0,...,0)'$. Under Assumption 3(*ii*), $\psi(\Lambda,\cdot)$ is smooth and strict monototic for Λ, we have $g(\Lambda)=g(\Lambda^0)+\sum_{l=1}^{\infty}\sum_{l_1+l_2+\cdots+l_P=l}\frac{\lambda_1^{l_1}\lambda_2^{l_2}\cdots\lambda_P^{l_P}}{l_1!l_2!\cdots l_P!}\cdot\frac{\partial^{l_1}\partial^{l_2}\cdots\partial^{l_P}g(\Lambda^0)}{\partial\lambda_1^{l_1}\partial\lambda_2^{l_2}\cdots\partial\lambda_P^{l_P}}$ where $0\leq l_1,...,l_P\leq l$. As $g(\Lambda^0)=\ln\left|I_n\right|=0$ (by Assumption 3(*ii*), $\psi(\Lambda^0,\cdot)\equiv 0$), it remains to consider derivatives. Notice that

16. Refer to Qu and Lee (2013).

$$\frac{\partial g(\Lambda)}{\partial \lambda_p} = -\text{tr}[\text{diag}(\frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_p})W_n S_n^{-1}(\Lambda)] = -\text{tr}[\text{diag}(\frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_p})W_n \sum_{j=0}^{\infty}(\Psi(\Lambda, Z_n)W_n)^j]$$
$$= -\text{tr}[\text{diag}(\frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_p})W_n \sum_{j=1}^{\infty}(\Psi(\Lambda, Z_n)W_n)^j], \ p = 1, .., P,$$

the third equality holds because $\text{tr}[\text{diag}(\frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_p})W_n] = 0$. For any $q, r = 1, ..., P$, we have

$$\frac{\partial^2 g(\Lambda)}{\partial \lambda_p \partial \lambda_q} = -\text{tr}[\text{diag}(\frac{\partial^2 \psi(\Lambda, z_{i,n})}{\partial \lambda_p \partial \lambda_q})W_n \sum_{j=1}^{\infty}(\Psi(\Lambda, Z_n)W_n)^j]$$
$$- \text{tr}[\text{diag}(\frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_p} \cdot \frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_q})W_n^2 \sum_{j=1}^{\infty} j(\Psi(\Lambda, Z_n)W_n)^{j-1}],$$

and

$$\frac{\partial^3 g(\Lambda)}{\partial \lambda_p \partial \lambda_q \partial \lambda_r} = -\text{tr}[\text{diag}(\frac{\partial^3 \psi(\Lambda, z_{i,n})}{\partial \lambda_p \partial \lambda_q \partial \lambda_r})W_n \sum_{j=1}^{\infty}(\Psi(\Lambda, Z_n)W_n)^j]$$
$$- \text{tr}[\text{diag}(\frac{\partial^2 \psi(\Lambda, z_{i,n})}{\partial \lambda_p \partial \lambda_q} \cdot \frac{\partial \psi(\Lambda, z_i)}{\partial \lambda_r})W_n^2 \sum_{j=1}^{\infty} j(\Psi(\Lambda, Z_n)W_n)^{j-1}]$$
$$- \text{tr}[\text{diag}(\frac{\partial^2 \psi(\Lambda, z_{i,n})}{\partial \lambda_p \partial \lambda_r} \cdot \frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_q} + \frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_p} \cdot \frac{\partial^2 \psi(\Lambda, z_{i,n})}{\partial \lambda_q \partial \lambda_r})W_n^2 \sum_{j=1}^{\infty} j(\Psi(\Lambda, Z_n)W_n)^{j-1}]$$
$$- \text{tr}[\text{diag}(\frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_p} \cdot \frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_q} \cdot \frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_r})W_n^3 \sum_{j=2}^{\infty} j(j-1)(\Psi(\Lambda, Z_n)W_n)^{j-2}].$$

At Λ^0,

$$\frac{\partial g(\Lambda^0)}{\partial \lambda_p} = 0,$$
$$\frac{\partial^2 g(\Lambda^0)}{\partial \lambda_p \partial \lambda_q} = -\sum_{i=1}^{n}[\frac{\partial \psi(\Lambda^0, z_{i,n})}{\partial \lambda_p}\frac{\partial \psi(\Lambda^0, z_{i,n})}{\partial \lambda_q}(W_n^2)_{ii}],$$
$$\frac{\partial^3 g(\Lambda^0)}{\partial \lambda_p \partial \lambda_q \partial \lambda_r} = -\sum_{i=1}^{n}[(\frac{\partial^2 \psi(\Lambda^0, z_i)}{\partial \lambda_p \partial \lambda_q}\frac{\partial \psi(\Lambda^0, z_{i,n})}{\partial \lambda_r} + \frac{\partial^2 \psi(\Lambda^0, z_{i,n})}{\partial \lambda_p \partial \lambda_r}\frac{\partial \psi(\Lambda^0, z_{i,n})}{\partial \lambda_q} + \frac{\partial \psi(\Lambda^0, z_{i,n})}{\partial \lambda_p}\frac{\partial^2 \psi(\Lambda^0, z_{i,n})}{\partial \lambda_q \partial \lambda_r})(W_n^2)_{ii}]$$
$$-\sum_{i=1}^{n}[\frac{\partial \psi(\Lambda^0, z_{i,n})}{\partial \lambda_p}\frac{\partial \psi(\Lambda^0, z_{i,n})}{\partial \lambda_q}\frac{\partial \psi(\Lambda^0, z_{i,n})}{\partial \lambda_r}(W_n^3)_{ii}].$$

By induction, $\frac{\partial^{l_1}\partial^{l_2}\cdots\partial^{l_P} g(\Lambda^0)}{\partial \lambda_1^{l_1}\partial \lambda_2^{l_2}\cdots\partial \lambda_P^{l_P}} = -\sum_{i=1}^{n}[\Xi_{i,n}^{(2)}(\Lambda^0, z_{i,n}, l_1, ..., l_P)(W_n^2)_{ii} + \cdots + \Xi_{i,n}^{(l)}(\Lambda^0, z_{i,n}, l_1, ..., l_P)(W_n^l)_{ii}]$, where $\Xi_{i,n}^{(h)}(\Lambda^0, z_{i,n}, l_1, ..., l_P), h = 1, ..., l$ are some combinations of the first, second, and higher order ($< l$) partial derivatives of $\psi(\Lambda, \cdot)$ and obviously

$$\Xi_{i,n}^{(l)}(\Lambda^0, z_{i,n}, l_1, ..., l_P) = \underbrace{\frac{\partial \psi(\Lambda^0, z_{i,n})}{\partial \lambda_1} \cdots \frac{\partial \psi(\Lambda^0, z_{i,n})}{\partial \lambda_1}}_{l_1} \cdots \underbrace{\frac{\partial \psi(\Lambda^0, z_{i,n})}{\partial \lambda_P} \cdots \frac{\partial \psi(\Lambda^0, z_{i,n})}{\partial \lambda_P}}_{l_P}.$$

Therefore, $\ln |I_n - \Psi(\Lambda, Z_n) W_n| = -\sum_{l=1}^{\infty} \sum_{l_1+l_2+\cdots+l_P=l} \frac{\lambda_1^{l_1} \lambda_2^{l_2} \cdots \lambda_P^{l_P}}{l_1! l_2! \cdots l_P!} \sum_{i=1}^{n} [\Xi_{i,n}^{(2)}(\Lambda^0, z_{i,n}, l_1, ..., l_P)(W_n^2)_{ii} + \cdots + \Xi_{i,n}^{(l)}(\Lambda^0, z_{i,n}, l_1, ..., l_P)(W_n^l)_{ii}]$.

First, we consider the pointwise convergence $\text{plim}_{n\to\infty} \frac{1}{n}(\ln |I_n - \Psi(\Lambda, Z_n) W_n| - E \ln |I_n - \Psi(\Lambda, Z_n) W_n|) = 0$, by the WLLN in Jenish and Prucha (2012), we only need to check the NED property and uniform L_η boundedness of $\ln |I_n - \Psi(\Lambda, Z_n) W_n|$. Based on the above Taylor expansion, we show that $\Xi_{i,n}^{(h)}(\Lambda^0, z_{i,n}, l_1, \ldots, l_P)(W_n^h)_{ii} \leq \tilde{c}_i^{*(h)} s^{-c_2 d_0}$ for some constant $\tilde{c}_i^{*(h)} > 0$. Given any distance s, the product terms in the summation $\sum_{j_1} \cdots \sum_{j_{h-1}}$ can be separated into two parts: the first part $\mathscr{P}(1)$, with the distance of each pair of successive nodes in the chain $i \to j_1 \to j_2 \to \cdots j_{h-1} \to i$ less than s/h, while the second part $\mathscr{P}(2)$ consists of the other product terms. Then in $\mathscr{P}(2)$, there exists at least one element among $\{w_{ij_1,n}, w_{j_1 j_2,n}, ..., w_{j_{h-1} i,n}\}$ that is $\leq c_1 (s/h)^{-c_2 d_0}$. Define W_{1n} as follows: $w_{ij,1,n} = w_{ij,n}$ when $w_{ij,n} \leq c_1 (s/h)^{-c_2 d_0}$; $w_{ij,1,n} = 0$ when $w_{ij,n} > c_1 (s/h)^{-c_2 d_0}$. W_{2n} is defined by $w_{ij,2,n} = w_{ij,n} - w_{ij,1,n}$. Thus every element in W_{2n} is either 0 or $> c_1 (s/h)^{-c_2 d_0}$. Hence,

$$\begin{aligned}
\sum_{\mathscr{P}(2)} w_{ij_1,n} w_{j_1 j_2,n} \ldots w_{j_{h-1} i,n} &\leq [(W_{1n} + W_{2n})^h]_{ii} - (W_{2n}^h)_{ii} \leq c_1 (s/h)^{-c_2 d_0} \sum_{k=0}^{h-1} \|W_{2n}\|_\infty^k \|W_n\|_1^{h-k-1} \\
&\leq \left[c_1 h^{c_2 d_0} \sum_{k=0}^{h-1} \|W_n\|_\infty^k \tilde{c}^* (h-k-1) c_w^{h-k-1} \right] s^{-c_2 d_0} \\
&\leq \left[\tilde{c}^* c_1 c_w^{h-1} h^{c_2 d_0} \sum_{k=0}^{h-1} (h-k-1) \right] s^{-c_2 d_0} \leq \left[\tilde{c}^* c_1 c_w^{h-1} h^{c_2 d_0 + 1} (h-1)/2 \right] s^{-c_2 d_0} \\
&= \tilde{c}_{2h}^* s^{-c_2 d_0} \text{ for some constant } \tilde{c}_{2h}^* > 0,
\end{aligned}$$

where the second and third inequalities follow from Lemma A.3 and Lemma 1 in Xu and Lee (2015b) respectively. Denote $g_{i,h} \equiv \Xi_{i,n}^{(h)}(\Lambda^0, z_{i,n}, l_1, \ldots, l_P)(W_n^h)_{ii}$, we have

$$\begin{aligned}
&\|g_{i,h} - \mathrm{E}[g_{i,h} | \mathscr{F}_{i,n}(s)]\|_2 \\
&\quad \leq \sum_{\mathscr{P}(1)} w_{ij_1,n} w_{j_1 j_2,n} \ldots w_{j_{h-1} i,n} \cdot \|\Xi_{i,n}^{(h)}\left(\Lambda^0, z_{i,n}, l_1, \ldots, l_P\right) - E[\Xi_{i,n}^{(h)}\left(\Lambda^0, z_{i,n}, l_1, \ldots, l_P\right) | \mathscr{F}_{i,n}(s)]\|_2 \\
&\quad + \sum_{\mathscr{P}(2)} w_{ij_1,n} w_{j_1 j_2,n} \ldots w_{j_{h-1} i,n} \\
&\quad \leq 2 c_w^h b_{\Xi_i^h} + \tilde{c}_{2h}^* s^{-c_2 d_0} \leq \tilde{c}_i^{*(h)} s^{-c_2 d_0} \text{ for some constant } \tilde{c}_i^{*(h)} > 0,
\end{aligned}$$

where the second inequality holds because the partial derivatives (and their combinations) in $\Xi_{i,n}^{(h)}(\Lambda^0, z_{i,n}, l_1, \ldots, l_P)$

are bounded, then $\Xi_{i,n}^{(h)}\left(\Lambda^0, z_{i,n}, l_1, \ldots, l_P\right)$ is bounded (suppose by $b_{\Xi_i^h}$) under Assumption 3(*ii*). The NED property of $\ln\left|I_n - \Psi\left(\Lambda, Z_n\right)W_n\right|$ follows from

$$
\begin{aligned}
&\|\ln\left|I_n - \Psi\left(\Lambda, Z_n\right)W_n\right| - E[\ln\left|I_n - \Psi\left(\Lambda, Z_n\right)W_n\right| |\mathcal{F}_{i,n}(s)]\|_2 \\
&\quad\leq \sum_{l=1}^{\infty}\sum_{l_1+\cdots+l_P=l}\frac{\lambda_1^{l_1}\cdots\lambda_P^{l_P}}{l_1!\cdots l_P!}\sum_{i=1}^{n}[\|g_{i,2}-E[g_{i,2}|\mathcal{F}_{i,n}(s)]\|_2+\cdots+\|g_{i,l}-E[g_{i,l}|\mathcal{F}_{i,n}(s)]\|_2] \\
&\quad\leq \sum_{l=1}^{\infty}\sum_{l_1+\cdots+l_P=l}\frac{\lambda_1^{l_1}\cdots\lambda_P^{l_P}}{l_1!\cdots l_P!}\sum_{i=1}^{n}\left(\tilde{c}_i^{*(2)}s^{-c_2d_0}+\cdots+\tilde{c}_i^{*(l)}s^{-c_2d_0}\right) \\
&\quad\leq s^{-c_2d_0}\sum_{l=1}^{\infty}\sum_{l_1+\cdots+l_{p_0}=l}\frac{\zeta_1^{l_1}\cdots\zeta_{p_0}^{l_{p_0}}}{l_1!\cdots l_{p_0}!}\sum_{i=1}^{n}(\tilde{c}_i^{*(2)}+\cdots+\tilde{c}_i^{*(l)}) \\
&\quad\leq \tilde{C}_0^* s^{-c_2d_0} \text{ for some constant } \tilde{C}_0^*>0.
\end{aligned}
$$

As

$$
\begin{aligned}
&|\ln|I_n-\Psi(\Lambda,Z_n)W_n|| \\
&=\left|\sum_{l=1}^{\infty}\sum_{l_1+\cdots+l_P=l}\frac{\lambda_1^{l_1}\cdots\lambda_P^{l_P}}{l_1!\cdots l_P!}\sum_{i=1}^{n}\left[\Xi_{i,n}^{(2)}(\Lambda^0,z_{i,n},l_1,...,l_P)(W_n^2)_{ii}+\cdots+\Xi_{i,n}^{(l)}(\Lambda^0,z_{i,n},l_1,...,l_P)(W_n^l)_{ii}\right]\right| \\
&\leq\sum_{l=1}^{\infty}\sum_{l_1+\cdots+l_P=l}\frac{\lambda_1^{l_1}\cdots\lambda_P^{l_P}}{l_1!\cdots l_P!}\sum_{i=1}^{n}\left[\left|\Xi_{i,n}^{(2)}(\Lambda^0,z_{i,n},l_1,...,l_P)\right|\|W_n\|_\infty^2+\cdots+\left|\Xi_{i,n}^{(l)}(\Lambda^0,z_{i,n},l_1,...,l_P)\right|\|W_n\|_\infty^l\right] \\
&\leq\sum_{l=1}^{\infty}\sum_{l_1+\cdots+l_P=l}\frac{\lambda_1^{l_1}\cdots\lambda_P^{l_P}}{l_1!\cdots l_P!}\sum_{i=1}^{n}(b_{\Xi_i^2}c_w^2+\cdots+b_{\Xi_i^l}c_w^l)<\infty,
\end{aligned}
$$

By Minkowski's inequality, we have the uniform L_η boundedness $\|\ln|I_n-\Psi(\Lambda,Z_n)W_n|\|_\eta<\infty$.

Second, it remains to check the stochastic equicontinuity of $\frac{1}{n}\ln|I_n-\Psi(\Lambda,Z_n)W_n|$. Applying the mean value theorem,

$$
\begin{aligned}
&\frac{1}{n}(\ln|I_n-\Psi(\Lambda_1,Z_n)W_n|-\ln|I_n-\Psi(\Lambda_2,Z_n)W_n|) \\
&=|(\lambda_{1,1}-\lambda_{1,2})\frac{1}{n}\mathrm{tr}(G_n(\bar{\lambda}_1))+\cdots+(\lambda_{P,1}-\lambda_{P,2})\frac{1}{n}\mathrm{tr}(G_n(\bar{\lambda}_P))| \\
&\leq|\lambda_{1,1}-\lambda_{1,2}|\cdot C_1^*+\cdots+|\lambda_{P,1}-\lambda_{P,2}|\cdot C_P^*
\end{aligned}
\tag{A.10}
$$

where $\bar{\Lambda}$ is between Λ_1 and Λ_2, $G_n(\lambda_p)=\mathrm{diag}(\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}W_n)$ and C_p^* is a constant for $p=1,..,P$. The inequality holds as $\sup_\Lambda\|G_n(\lambda_p)\|_\infty<\infty$.

Proof of Theorem 1. With the identification condition in Assumption 6, it suffices to show the consistency of the QMLE in two steps:

Step 1: Uniform convergence $\frac{1}{n}[\sup_{\theta\in\Theta}|\ln L_n(\theta)-E\ln L_n(\theta)|]\xrightarrow{p}0$. Denote $\omega=(\Lambda',\beta',\mathrm{vec}(\Gamma)',\delta')$

and $v_{i,n}(\omega) = y_{i,n} - \psi(\Lambda, z_{i,n})w_{i\cdot,n}Y_n - x'_{1,in}\beta - (z_{i,n} - x_{2,in}\Gamma)'\delta = [\psi(\Lambda_0, z_{i,n}) - \psi(\Lambda, z_{i,n})]w_{i\cdot,n}Y_n + x'_{1,in}(\beta_0 - \beta) + x'_{2,in}(\Gamma - \Gamma_0) + \varepsilon'_{i,n}(\delta_0 - \delta) + \xi_{i,n}$. With Proposition 2, it remains to show the uniform convergence $\text{plim}_{n\to\infty}\text{sup}_{\theta\in\Theta}\frac{1}{n}|\sum_{i=1}^n v_{i,n}^2(\omega) - Ev_{i,n}^2(\omega)| = 0$ by showing the pointwise convergence $\text{plim}_{n\to\infty}\frac{1}{n}|\sum_{i=1}^n v_{i,n}^2(\omega) - Ev_{i,n}^2(\omega)| = 0$ for each ω, and the stochastic equicontinuity of $v_{i,n}^2(\omega)$. The pointwise convergence holds by WLLN in Jenish and Prucha (2012) because with Proposition 1 and under Assumptions 2 and 7(*ii*), $v_{i,n}(\omega)$ is L_5 bounded uniformly in i and n, and uniformly L_2-NED, then $v_{i,n}^2(\omega)$ is $L_{2.5}$ bounded and uniformly L_2-NED by Claim B.3 in Qu and Lee (2015). By Lemma 1 in Andrews (1992), the stochastic equicontinuity follows from the uniform L_2 boundedness of $\psi(\Lambda, z_{i,n})w_{i\cdot,n}Y_n$, $x_{i,n}$, $\varepsilon_{i,n}$, and

$$
\begin{aligned}
&|\frac{1}{n}\sum_{i=1}^n v_{i,n}^2(\omega_1) - \frac{1}{n}\sum_{i=1}^n v_{i,n}^2(\omega_2)| \\
&= |\frac{1}{n}\sum_{i=1}^n [v_{i,n}(\omega_1) + v_{i,n}(\omega_2)] \cdot [(\psi(\Lambda_2, z_{i,n}) - \psi(\Lambda_1, z_{i,n}))w_{i\cdot,n}Y_n + x'_{1,in}(\beta_2 - \beta_1) + x'_{2,in}(\Gamma_2 - \Gamma_1) + \varepsilon'_{i,n}(\delta_2 - \delta_1)]| \\
&\leq \frac{1}{n}\sum_{i=1}^n [4b_\psi|w_{i\cdot,n}Y_n| + 4\sum_{\iota_1=1}^{k_1}|x_{1,i\iota_1,n}|\text{sup}_{\beta_{\iota_1}}|\beta_{\iota_1}| + 4\sum_{\iota_2=1}^{k_2}|x_{2,i\iota_2,n}|\text{sup}_{\Gamma_{\iota_2}}|\Gamma_{\iota_2}| + 4\sum_{l=1}^{h}|\varepsilon_{il,n}|\text{sup}_{\delta_l}|\delta_l| \\
&\qquad + 2|\xi_{i,n}|] \cdot [|\psi(\Lambda_2, z_{i,n}) - \psi(\Lambda_1, z_{i,n})||w_{i\cdot,n}Y_n| + \sum_{\iota_1=1}^{k_1}|x_{1,i\iota_1,n}||\beta_{2\iota_1} - \beta_{1\iota_1}| \\
&\qquad + \sum_{\iota_2=1}^{k_2}|x_{2,i\iota_2,n}||\Gamma_{2\iota_2} - \Gamma_{1\iota_1}| + \sum_{l=1}^{h}|\varepsilon_{il,n}||\delta_{2l} - \delta_{1l}|].
\end{aligned}
$$

Step 2: Uniform equicontinuity of $\lim_{n\to\infty}E(\frac{1}{n}\ln(\theta))$. With the stochastic equicontinuity of $v_{i,n}^2(\omega)$ and the boundedness of the parameter space, by Corollary 3.1 in Newey (1991), we have the equicontinuity of $\sigma_\xi^{-2}E[S_n(\Lambda) - X_{1n}\beta - (Z_n - X_{2n}\Gamma)\delta]'[S_n(\Lambda) - X_{1n}\beta - (Z_n - X_{2n}\Gamma)\delta]$ by the inequality in (A.10), variance parameters are bounded away from zero in compact parameter space and by the proof in Proposition 2, $\frac{1}{n}E|\ln(I_n - \Psi(\Lambda, Z_n)W_n)| = O(1)$, then $E(\frac{1}{n}\ln(\theta))$ is uniformly equicontinuous in $\theta \in \Theta$.

Proof of Theorem 2. By Taylor expansion, $\sqrt{n}(\hat{\theta} - \theta_0) = (-\frac{1}{n}\frac{\partial^2 \ln L_n(\bar{\theta})}{\partial\theta\partial\theta'})^{-1}(\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta})$, where $\bar{\theta}$ lies between θ_0 and $\hat{\theta}$. To derive the asymptotic distribution of the extremum estimator, we need to show the asymptotic normality of the consistency root of $\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\hat{\theta})}{\partial\theta} = 0$ and $\frac{1}{n}|\frac{\partial^2 \ln L_n(\hat{\theta})}{\partial\theta\partial\theta'} - \text{E}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}| \xrightarrow{p} 0$.

Step 1: Asymptotic normality of the score vector. Denote $r_{p,i,n} = \frac{\partial\psi(\Lambda_0, z_{i,n})}{\partial\lambda_p}\sum_{j=1}^n w_{ij,n}y_{j,n}$, $g_{p,ii,n} = \frac{\partial\psi(\Lambda_0, z_{i,n})}{\partial\lambda_p}\sum_{l=1}^\infty [W_n(\Psi(\Lambda_0, Z_n)W_n)^l]_{ii}$, $\kappa(\Sigma_{\varepsilon,0}) = [\text{tr}(\Sigma_{\varepsilon,0}^{-1}\frac{\partial\Sigma_{\varepsilon,0}}{\partial\alpha_1}), \cdots, \text{tr}(\Sigma_{\varepsilon,0}^{-1}\frac{\partial\Sigma_{\varepsilon,0}}{\partial\alpha_J})]'$. From the first order conditions, we can write the score as a summation:

$$\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} r_{1,i,n}\xi_{i,n}/\sigma_{\xi,0}^2 - g_{1,ii,n} - \mathrm{E}[r_{1,i,n}\xi_{i,n}/\sigma_{\xi,0}^2 - g_{1,ii,n}] \\ \vdots \\ r_{P,i,n}\xi_{i,n}/\sigma_{\xi,0}^2 - g_{P,ii,n} - \mathrm{E}[r_{P,i,n}\xi_{i,n}/\sigma_{\xi,0}^2 - g_{P,ii,n}] \\ x_{1,in}\xi_{i,n}/\sigma_{\xi,0}^2 \\ (\Sigma_{\varepsilon,0}^{-1}\otimes x_{2,in})(z_{i,n}-\Gamma_0' x_{2,in}) - \frac{1}{\sigma_{\xi,0}^2}\cdot \delta_0 \otimes x_{2,in}\xi_{i,n} \\ (\xi_{i,n}^2 - \sigma_{\xi,0}^2)/2\sigma_{\xi,0}^4 \\ \varepsilon_{i,n}\xi_{i,n}/2\sigma_{\xi,0}^2 \\ -\frac{1}{2}\kappa(\Sigma_{\varepsilon,0}) - \frac{1}{2}\frac{\partial}{\partial\alpha}\mathrm{tr}[\Sigma_{\varepsilon,0}^{-1}\varepsilon_{i,n}\varepsilon_{i,n}'] \end{pmatrix}.$$

Denote $q_{i,n} \equiv [(r_{1,i,n}\xi_{i,n}/\sigma_{\xi,0}^2 - g_{1,ii,n} - \mathrm{E}[r_{1,i,n}\xi_{i,n}/\sigma_{\xi,0}^2 - g_{1,ii,n}])^2 + \cdots + (r_{P,i,n}\xi_{i,n}/\sigma_{\xi,0}^2 - g_{P,ii,n} - \mathrm{E}[r_{P,i,n}\xi_{i,n}/\sigma_{\xi,0}^2 - g_{P,ii,n}])^2 + (x_{1,in}\xi_{i,n}/\sigma_{\xi,0}^2)^2 + ((\Sigma_{\varepsilon,0}^{-1}\otimes x_{2,in})(z_{i,n}-\Gamma_0'x_{2,in}) - \frac{1}{\sigma_{\xi,0}^2}\cdot\delta_0\otimes x_{2,in}\xi_{i,n})^2 + ((\xi_{i,n}^2-\sigma_{\xi,0}^2)/2\sigma_{\xi,0}^4)^2 + (\varepsilon_{i,n}\xi_{i,n}/2\sigma_{\xi,0}^2)^2 + (\frac{1}{2}\kappa(\Sigma_{\varepsilon,0}) - \frac{1}{2}\frac{\partial}{\partial\alpha}\mathrm{tr}[\Sigma_{\varepsilon,0}^{-1}\varepsilon_{i,n}\varepsilon_{i,n}'])^2]^{\frac{1}{2}}$. First, we show the uniform $L_{2+\phi_1}$ integrability for some $\phi_1 > 0$. One sufficient condition from page 54 in Shorack (2000) is to show $\sup_{i,n}\mathrm{E}q_{i,n}^{2+\phi_2} < \infty$ for some $\phi_2 > 0$. We have

$$\begin{aligned}\sup_{i,n}\mathrm{E}q_{i,n}^{2.5} \leq \sup_{i,n}&(P+k_1+k_2h+h+1)^{1.5} \\ &\cdot[\mathrm{E}|r_{1,i,n}\xi_{i,n}/\sigma_{\xi,0}^2 - g_{1,ii,n} - \mathrm{E}[r_{1,i,n}\xi_{i,n}/\sigma_{\xi,0}^2 - g_{1,ii,n}]|^2 + \cdots \\ &+\mathrm{E}|(r_{P,i,n}\xi_{i,n}/\sigma_{\xi,0}^2 - g_{P,ii,n} - \mathrm{E}[r_{P,i,n}\xi_{i,n}/\sigma_{\xi,0}^2 - g_{P,ii,n}]|^2 \\ &+\sum_{\iota_1=1}^{k_1}\mathrm{E}|\frac{x_{1,i\iota_1,n}\xi_{i,n}}{\sigma_{\xi,0}^2}|^2 + \sum_{\iota_2=1}^{k_2h}\mathrm{E}|\{(\Sigma_{\varepsilon,0}^{-1}\otimes x_{2,i,n})(z_{i,n}-\Gamma_0'x_{2,in}) - \frac{1}{\sigma_{\xi,0}^2}\cdot\delta_0\otimes x_{2,in}\xi_{i,n}\}_{\iota_2\times 1}|^2 \\ &+\mathrm{E}|(\xi_{i,n}^2-\sigma_{\xi,0}^2)/2\sigma_{\xi,0}^4|^2 + \sum_{\tau=1}^{h}\mathrm{E}|\varepsilon_{i\tau,n}\xi_{i,n}/2\sigma_{\xi,0}^2|^2 \\ &+\mathrm{E}|\frac{1}{2}\kappa(\Sigma_{\varepsilon,0}) - \frac{1}{2}\frac{\partial}{\partial\alpha}\mathrm{tr}[\Sigma_{\varepsilon,0}^{-1}\varepsilon_{i,n}\varepsilon_{i,n}']|^2] < \infty,\end{aligned}$$

which holds as $\{r_{p,i,n}\xi_{i,n}/\sigma_{\xi,0}^2\}$ is uniformly $L_{2.5}$ bounded by Proposition 1 and $\sup_{i,n}|\mathrm{E}r_{p,i,n}\xi_{i,n}| \leq \sup_{i,n}\mathrm{E}|r_{p,i,n}\xi_{i,n}| \leq \sup_{i,n}\|r_{p,i,n}\xi_{i,n}\|_p < \infty$, $\{\xi_{i,n}\}$, $\{\varepsilon_{i,n}\}$, $\{x_{i,n}\}$ are uniformly L_5 bounded by Assumption 7(ii), and for $p = 1, .., P$,

$$
\begin{aligned}
&|g_{p,ii,n} - \mathrm{E} g_{p,ii,n}| \\
&\leq 2 \sum_{l=0}^{\infty} |\frac{\partial \psi(\Lambda_0, z_{i,n})}{\partial \lambda_p}| \cdot b_\psi^l \sum_{j_1} \sum_{j_2} \cdots \sum_{j_l} w_{ij_1,n} w_{j_1 j_2,n} \cdots w_{j_{l-1} j_l,n} w_{j_l i,n} \\
&\leq 2 \sum_{l=0}^{\infty} |\frac{\partial \psi(\Lambda_0, z_{i,n})}{\partial \lambda_p}| b_\psi^l c_w^{l+1} \leq \frac{2 c_w |\frac{\partial \psi(\Lambda_0, z_{i,n})}{\partial \lambda_p}|}{1 - b_\psi c_w} < \infty.
\end{aligned}
$$

Second, we can show the L_2-NED property of $\{q_{i,n}\}$ on the i.i.d. random fields $\varsigma_{i,n} = (\varepsilon'_{i,n}, \xi_{i,n})'$. For $\{g_{p,ii,n}\}_{p=1}^P$, similar to the proof of Proposition 2, we have

$$
\begin{aligned}
&\|g_{p,ii,n} - E\left[g_{p,ii,n} | \mathscr{F}_{i,n}(s)\right]\|_2 \\
&\leq |\frac{\partial \psi(\Lambda_0, z_{i,n})}{\partial \lambda_p}| \sum_{l=1}^{\infty} \sum_{\mathscr{P}(1)} w_{ik_1,n} w_{k_1 k_2,n} \cdots w_{k_{l-1} k_l,n} w_{k_l i,n} \cdot \|(\Psi^l(\Lambda_0, Z_n))_{ii} - E[(\Psi^l(\Lambda_0, Z_n))_{ii} | \mathscr{F}_{i,n}(s)]\|_2 \\
&\quad + |\frac{\partial \psi(\Lambda_0, z_{i,n})}{\partial \lambda_p}| \sum_{l=1}^{\infty} \sum_{\mathscr{P}(2)} w_{ik_1,n} w_{k_1 k_2,n} \cdots w_{k_{l-1} k_l,n} w_{k_l i,n} \\
&\leq 2 |\frac{\partial \psi(\Lambda_0, z_{i,n})}{\partial \lambda_p}| \sum_{l=1}^{\infty} c_w^{l+1} b_\psi^l + |\frac{\partial \psi(\Lambda_0, z_{i,n})}{\partial \lambda_p}| \sum_{l=1}^{\infty} \tilde{c}^*_{2(l+1)} s^{-c_2 d_0} \leq c_7 s^{-c_2 d_0}
\end{aligned}
$$

for some constant $c_7 > 0$. The L_2-NED property of $\{q_{i,n}\}$ is a result of Lemma A.3 in Xu and Lee (2015a). As the error terms are i.i.d and $\{\varepsilon_{i,n}, \xi_{i,n}, x_{i,n}\}_{i=1}^n$ satisfies Assumption 8, then Assumption 3 in Jenish and Prucha (2012) is satisfied. Besides, Condition (b) in Assumption 4 in Jenish and Prucha (2012) is met under Assumption 9. Then all the conditions for the CLT of the L_2-NED sequence are satisfied and we have $\frac{1}{\sqrt{n}} \sum_{i=1}^n (r_{1,i,n} \xi_{i,n} / \sigma_{\xi,0}^2 - g_{1,ii,n} - E[r_{1,i,n} \xi_{i,n} / \sigma_{\xi,0}^2 - g_{1,ii,n}], \cdots, r_{P,i,n} \xi_{i,n} / \sigma_{\xi,0}^2 - g_{P,ii,n} - E[r_{P,i,n} \xi_{i,n} / \sigma_{\xi,0}^2 - g_{P,ii,n}], x_{1,in} \xi_{i,n} / \sigma_{\xi,0}^2, (\Sigma_{\varepsilon,0}^{-1} \otimes x_{2,in})(z_{i,n} - \Gamma_0' x_{2,in}) - \frac{1}{\sigma_{\xi,0}^2} \cdot \delta_0 \otimes x_{2,in} \xi_{i,n}, (\xi_{i,n}^2 - \sigma_{\xi,0}^2)/2\sigma_{\xi,0}^4, \varepsilon_{i,n} \xi_{i,n} / 2\sigma_{\xi,0}^2, -\frac{1}{2}\kappa(\Sigma_{\varepsilon,0}) - \frac{1}{2} \frac{\partial}{\partial \alpha} \mathrm{tr}[\Sigma_{\varepsilon,0}^{-1} \varepsilon_{i,n} \varepsilon'_{i,n}])' \xrightarrow{d} N(0, \Sigma_0)$.

Step 2. The uniform convergence of the Hessian matrix. The second order derivatives are provided in Appendix A. First, we consider $\frac{1}{n} |\frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta \partial \theta'} - \mathrm{E} \frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta \partial \theta'}| \xrightarrow{p} 0$. With the L_5 boundedness of $\{\xi_{i,n}, \varepsilon_{i,n}, x_{i,n}\}_{i=1}^n$, $\{\psi(\Lambda_0, z_{i,n}) w_{i\cdot,n} Y_n\}_{i=1}^n$, and $\{v_{i,n}(\omega)\}_{i=1}^n$ uniformly in i and n, and their uniform L_2-NED properties, their products satisfy the weak LLN in Jenish and Prucha (2012). Then it remains to show that $\frac{1}{n}\{\mathrm{tr}[\mathrm{diag}(\frac{\partial^2 \psi(\Lambda, z_{i,n})}{\partial \lambda_p \partial \lambda_1}) W_n S_n^{-1}(\Lambda)] - E\mathrm{tr}[\mathrm{diag}(\frac{\partial^2 \psi(\Lambda, z_{i,n})}{\partial \lambda_p \partial \lambda_1}) W_n S_n^{-1}(\Lambda)]\} \xrightarrow{p} 0$ and $\frac{1}{n}\{\mathrm{tr}[\mathrm{diag}(\frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_p} \cdot \frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_q})(W_n S_n^{-1}(\Lambda))^2] - E\mathrm{tr}[\mathrm{diag}(\frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_p} \cdot \frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_q})(W_n S_n^{-1}(\Lambda))^2]\} \xrightarrow{p} 0$, $p, q = 1, ..., P$, which can be done by showing that $\{\frac{\partial^2 \psi(\Lambda, z_{i,n})}{\partial \lambda_p \partial \lambda_q}(W_n S_n^{-1}(\Lambda))_{ii}\}$ and $\{\frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_p} \cdot \frac{\partial \psi(\Lambda, z_{i,n})}{\partial \lambda_q}([W_n S_n^{-1}(\Lambda)]^2)_{ii}\}$ are uniformly bounded and L_2-NED uniformly in i and n. The uniform boundedness holds because $|\frac{\partial^2\psi(\Lambda,z_{i,n})}{\partial\lambda_p\partial\lambda_q}(W_nS_n^{-1}(\Lambda))_{ii}| \leq \sum_{l=0}^{\infty}|\frac{\partial^2\psi(\Lambda,z_{i,n})}{\partial\lambda_p\partial\lambda_q}|b_\psi^l c_w^{l+1} \leq \frac{c_w|\frac{\partial^2\psi(\Lambda,z_{i,n})}{\partial\lambda_p\partial\lambda_q}|}{1-b_\psi c_w}$, and

$$
\begin{aligned}
|\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}([W_nS_n^{-1}(\Lambda)]^2)_{ii}| &\leq |\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}|\sum_{l=0}^{\infty}b_\psi^l c_w^{l+1}\sum_{l'=0}^{\infty}b_\psi^{l'}c_w^{l'+1} \\
&= |\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}|\sum_{k=0}^{\infty}(1+k)b_\psi^{k+2}c_w^{k+2} < \infty.
\end{aligned}
$$

Under Assumption 3(*ii*), the uniform L_2-NED property of $\{\frac{\partial^2\psi(\Lambda,z_{i,n})}{\partial\lambda_p\partial\lambda_q}(W_nS_n^{-1}(\Lambda))_{ii}\}$ can be shown similar to the proof of the uniform L_2-NED of $g_{p,ii,n}$ in Step 1. For $\{\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}([W_nS_n^{-1}(\Lambda)]^2)_{ii}\}$, since $([W_nS_n^{-1}(\Lambda)]^2)_{ii} = \left(\sum_{l=0}^{\infty}W_n(\Psi(\Lambda,Z_n)W_n)^l\sum_{l'=0}^{\infty}W_n(\Psi(\Lambda,Z_n)W_n)^{l'}\right)_{ii}$, we have

$$
\begin{aligned}
&\left\|\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}\left([W_nS_n^{-1}(\Lambda)]^2\right)_{ii} - E\left[\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}\left([W_nS_n^{-1}(\Lambda)]^2\right)_{ii}|\mathcal{F}_{i,n}(s)\right]\right\|_2 \\
&\leq \left|\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}\right|\sum_{k=0}^{\infty}\sum_{l=0}^{k}\sum_{\mathcal{P}(1)}w_{ij_1,n}\cdots w_{j_kj_{k+1},n}w_{j_{k+1}i,n}\left\|\left(\Psi^k\left(\Lambda_0,Z_n\right)\right)_{ii} - \mathrm{E}\left[\left(\Psi^k\left(\Lambda_0,Z_n\right)\right)_{ii}\mid\mathcal{F}_{i,n}(s)\right]\right\|_2 \\
&\quad + \left|\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}\right|\sum_{k=0}^{\infty}\sum_{l=0}^{k}\sum_{\mathcal{P}(2)}w_{ij_1,n}\cdots w_{j_kj_{k+1},n}w_{j_{k+1}i,n} \\
&\leq 2\left|\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}\right|\sum_{k=0}^{\infty}(1+k)b_\psi^k c_w^{k+1} + \left|\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}\right|\sum_{k=0}^{\infty}(1+k)\tilde{c}_{2(k+1)}^{*}s^{-c_2d_0} \\
&\leq c_8 s^{-c_2d_0} \text{ for some constant } c_8 > 0.
\end{aligned}
$$

Second, we consider $\frac{1}{n}|\frac{\partial^2\ln L_n(\hat\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\ln L_n(\theta_0)}{\partial\theta\partial\theta'}| \xrightarrow{p} 0$. By Theorem 1, $\hat\theta-\theta_0 \xrightarrow{p} 0$, all the other terms can be checked easily except $\frac{\partial^2\ln L_n(\theta)}{\partial\lambda_p\partial\lambda_q}, p,q = 1,...,P$. It suffices to show that $\frac{\partial}{\partial\lambda_r}\frac{1}{n}\mathrm{tr}[\mathrm{diag}(\frac{\partial^2\psi(\Lambda,z_{i,n})}{\partial\lambda_p\partial\lambda_q})W_nS_n^{-1}(\Lambda)] = \frac{1}{n}\mathrm{tr}[\mathrm{diag}(\frac{\partial^3\psi(\Lambda,z_{i,n})}{\partial\lambda_p\partial\lambda_q\partial\lambda_r})W_nS_n^{-1}(\Lambda)] + \frac{1}{n}\mathrm{tr}[\mathrm{diag}(\frac{\partial^2\psi(\Lambda,z_{i,n})}{\partial\lambda_p\partial\lambda_q}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_r})(W_nS_n^{-1}(\Lambda))^2]$ and $\frac{\partial}{\partial\lambda_r}\frac{1}{n}\mathrm{tr}[\mathrm{diag}(\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q})(W_nS_n^{-1}(\Lambda))^2] = \frac{1}{n}\mathrm{tr}[\mathrm{diag}(\frac{\partial^2\psi(\Lambda,z_{i,n})}{\partial\lambda_p\partial\lambda_r}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q} + \frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial^2\psi(\Lambda,z_{i,n})}{\partial\lambda_q\partial\lambda_r})(W_nS_n^{-1}(\Lambda))^2] + \frac{2}{n}\mathrm{tr}[\mathrm{diag}(\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_r})(W_nS_n^{-1}(\Lambda))^3]$ are bounded for $r = 1,..,P$, which can be done by showing that

$$
\begin{aligned}
&(\frac{\partial^3\psi(\Lambda,z_{i,n})}{\partial\lambda_p\partial\lambda_q\partial\lambda_r})[W_nS_n^{-1}(\Lambda)]_{ii},\ (\frac{\partial^2\psi(\Lambda,z_{i,n})}{\partial\lambda_p\partial\lambda_q}\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_r})[(W_nS_n^{-1}(\Lambda))^2]_{ii}, \\
&(\frac{\partial^2\psi(\Lambda,z_i)}{\partial\lambda_p\partial\lambda_r}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q} + \frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial^2\psi(\Lambda,z_{i,n})}{\partial\lambda_q\partial\lambda_r})[(W_nS_n^{-1}(\Lambda))^2]_{ii}, \\
&(\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_r})[W_nS_n^{-1}(\Lambda))^3]_{ii}
\end{aligned}
$$

are bounded, the first three terms can be shown as above for the uniform boundedness, and the fourth one holds because

$$
\begin{aligned}
&|(\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_r})[W_nS_n^{-1}(\Lambda))^3]_{ii}| \\
&\leq |\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}\cdot\frac{\partial\psi(\Lambda,z_i)}{\partial\lambda_r}|\sum_{l=0}^{\infty}b_\psi^l c_w^{l+1}\sum_{l'=0}^{\infty}b_\psi^{l'}c_w^{l'+1}\sum_{l''=0}^{\infty}b_\psi^{l''}c_w^{l''+1} \\
&= |\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_q}\cdot\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_r}|\sum_{k=0}^{\infty}\sum_{l+l'+l''=k}b_\psi^k c_w^{k+3} \\
&= |\frac{\partial\psi(\Lambda,z_{i,n})}{\partial\lambda_p}\frac{\partial\psi(\Lambda,z_i)}{\partial\lambda_q}\frac{\partial\psi(\Lambda,z_i)}{\partial\lambda_r}||b_\psi^{-3}|\sum_{k=0}^{\infty}\frac{1}{2}(k+1)(k+2)(b_\psi c_2)^{k+2}<\infty
\end{aligned}
$$

The last inequality hods as under Assumption 3(*ii*), $b_\psi c_w < 1$.

Therefore, we have $\sqrt{n}\left(\hat{\theta}-\theta_0\right)\xrightarrow{d}N\left(0,\Sigma_{QMLE}\right)$, where $\Sigma_{QMLE}=(\lim_{n\to\infty}\frac{1}{n}\mathrm{E}(\frac{\partial^2\ln L_n(\theta_0)}{\partial\theta\partial\theta'}))^{-1}\times\lim_{n\to\infty}\frac{1}{n}\mathrm{E}(\frac{\partial\ln L_n(\theta_0)}{\partial\theta}\frac{\partial\ln L_n(\theta_0)}{\partial\theta'})(\lim_{n\to\infty}\frac{1}{n}\mathrm{E}(\frac{\partial^2\ln L_n(\theta_0)}{\partial\theta\partial\theta'}))^{-1}$. The expressions for each term of Σ_{QMLE} are provided in Appendix A.

References

Akaike, H. 1973. “Information theory and an extension of the maximum likelihood principle.” *Petrov, B.N., Csáki, F., (eds.), 2nd International Symposium on Information Theory,* 267–281.

Andrews, D. 1992. “Estimation of a SAR model with endogenous spatial weights constructed by bilateral variables.” *Econometric Theory* 8 (2): 241–257.

Brock, W., and S. Durlauf. 2001. “Discrete Choice with Social Interactions.” *The Review of Economic Studies* 68 (2): 235–260.

Dong, Y. 2010. “The Spillover Effect of Education on Productivity: Evidence from China.” 2010 International Conference on e-Education, e-Business, e-Management and e-Learning, pp. 420-424.

Gennaioli, N., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer. 2013. “Human Capital and Regional Development.” *The Quarterly Journal of Economics* 128 (1): 105–164.

Hoshino, T. 2022. “Sieve IV estimation of cross-sectional interaction models with nonparametric endogenous effect.” *Journal of Econometrics* 229 (2): 263–275.

Jenish, N., and I. Prucha. 2009. “Central limit theorems and uniform laws of large numbers for arrays of random fields.” *Journal of Econometrics* 150 (1): 86–98.

Jenish, N., and I. Prucha. 2012. “On spatial processes and asymptotic inference under near-epoch dependence.” *Journal of Econometrics* 170 (1): 178–190.

Jeong, H., Y. Lin, and L.-f. Lee. 2021. “Maximum likelihood estimation of a spatial autoregressive model for origin-destination flow variables.” Working Paper, https://drive.google.com/file/d/1CJVtwUaM4AziIG0twSs_0QZ7q7lUfHmZ/view, November.

Malikov, E., and Y. Sun. 2017. “Semiparametric Estimation and Testing of Smooth Coefficient Spatial Autoregressive Models.” *Journal of Econometrics* 199 (1): 12–34.

Martin, R., and P. Sunley. 1998. “Slow Convergence? The New Endogenous Growth Theory and Regional Development.” *Economic Geography* 74 (3): 201–227.

Moretti, E. 2004. “Workers’ Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions.” *American Economic Review* 94 (3): 656–690.

Newey, W. 1991. “Uniform convergence in probability and stochastic equicontinuity.” *Econometrica* 59 (4): 1161–1167.

Paelinck, J., and L. Klaassen. 1979. *Spatial Econometrics.* Farnborough: Saxon House.

Palivos, T., and P. Wang. 1996. “Spatial Agglomeration and Endogenous Growth.” *Regional Science and Urban Economics* 26 (6): 645–669.

Qu, X., and L. F. Lee. 2013. “Locally most powerful tests for spatial interactions in the simultaneous SAR Tobit model.” *Regional Science and Urban Economics* 43 (2): 307–321.

———. 2015. “Estimating a Spatial Autoregressive model with an Endogenous Spatial Weight Matrix.” *Journal of Econometrics* 184 (2): 209–232.

Qu, X., L. F. Lee, and C. Yang. 2021. “Estimation of a SAR model with endogenous spatial weights constructed by bilateral variables.” *Journal of Econometrics* 221 (2): 180–197.

Qu, X., L. F. Lee, and J. Yu. 2017. “QML Estimation of Spatial Dynamic Panel Data Models with Endogenous Time Varying Spatial Weights Matrices.” *Journal of Econometrics* 197 (2): 173–201.

Rothenberg, T. 1971. “Identification in parametric models.” *Econometrica* 39:577–591.

Sanso-Navarro, M., M. Vera-Cabello, and D. P. Ximénez-de-Embún. 2017. “Human Capital Spillovers and Regional Development.” *Journal of Applied Econometrics* 32 (4): 923–930.

Shorack, G. 2000. *Probability for Statisticians.* New York: Springer-Verlag.

White, H. 1982. “Maximum Likelihood Estimation of Misspecified Models.” *Econometrica* 50 (1): 1–25.

Xu, X., and L. Lee. 2015a. “A spatial autoregressive model with a nonlinear transformation of the dependent variable.” *Journal of Econometrics* 186 (1): 1–18.

———. 2015b. “Maximum likelihood estimation of a spatial autoregressive Tobit model.” *Journal of Econometrics* 188 (1): 264–280.