SUPPLEMENTARY APPEXNDIX
TO THE PAPER
ADDRESSING ENDOGENEITY ISSUES IN A SAR MODEL USING COPULAS

YANLI LIN* and YICHUN SONG†

October 15, 2022

S.1 Direct Extension of Park and Gupta (2012)'s approach

In this section, we consider a direct extension of Park and Gupta (2012)'s approach to a SAR model and show the additional assumptions (Assumptions S.1-S.2) needed to guarantee the consistency of the estimation methods proposed in the main draft.

Assumption S.1. *x_i and z_i^* are uncorrelated.*

Assumption S.2. *Non-normality of $z_{\iota,i}(\iota = 1, ..., p)$ when z_i are added as endogenous explanatory variables.*

S.1.1 Copula modelling

Unlike the approaches in Haschka (2022) and Yang et al. (2022), Park and Gupta (2012) only model the correlations between the endogenous variables z_i and the error term v_i using a parametric copula. Given the known marginal distributions of $z_{1,i}, ..., z_{p,i}$ and v_i, denoted as $H_1(z_{1,i}), ..., H_p(z_{p,i})$ and $G(v_i)$, by Sklar's theorem (Sklar, 1959), we can construct a multivariate joint distribution,

$$F(v_i, z_{1,i}, ..., z_{p,i}) = C(U_{v_i}, U_{z_{1,i}}, ..., U_{z_{p,i}}) \tag{S.1}$$

where $C(\cdot) : [0, 1]^{p+1} \to [0, 1]$ is a $p+1$-dimensional copula function, $U_{v_i} = G(v_i), U_{z_{1,i}} = H_1(z_{1,i}), ..., U_{z_{p,i}} = H_p(z_{p,i})$. The joint density function is

$$f(v_i, z_{1,i}, ..., z_{p,i}) = c(U_{v_i}, U_{z_{1,i}}, ..., U_{z_{p,i}})g(v_i)\prod_{\iota=1}^{p} h_\iota(z_{\iota,i}) \tag{S.2}$$

where $c(U_{v_i}, U_{z_{1,i}}, ..., U_{z_{p,i}}) = \frac{\partial^{p+1}C(U_{v_i}, U_{z_{1,i}}, ..., U_{z_{p,i}})}{\partial v_i \partial z_{1,i} \cdots \partial z_{p,i}}$, $g(\cdot)$ and $h_\iota(\cdot)(\iota = 1, ..., p)$ are marginal density functions of $G(\cdot)$ and $H_\iota(\cdot)(\iota = 1, ..., p)$ respectively. As in Park and Gupta (2012), we consider

*Department of Economics, The Ohio State University, 1945 N. High Street, Columbus, Ohio. Email: lin.3019@osu.edu.

†Department of Economics, The Ohio State University, 1945 N. High Street, Columbus, Ohio. Email: song.1399@osu.edu.

the Gaussian copula. The $(p+1)$- dimensional Gaussian copula with a correlation matrix $P \in [-1,1]^{(p+1)\times(p+1)}$, where $P = \begin{pmatrix} 1 & \rho'_{vz} \\ \rho_{vz} & P_z \end{pmatrix}$ with $\rho_{vz} = (\rho_{vz_1,\dots,\rho_{vz_p}})'$ and $P_z = \begin{pmatrix} 1 & \rho_{z_1z_2} & \cdots & \rho_{z_1z_p} \\ * & 1 & \cdots & \rho_{z_2z_p} \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & 1 \end{pmatrix}$, can be written as

$$C_P^{Gaussian}(U_{v_i}, U_{z_{1,i}}, \dots, U_{z_{p,i}}) = \Phi_P(v_i^*, z_{1,i}^*, \dots, z_{p,i}^*) \tag{S.3}$$

where $v_i^* = \Phi^{-1}(U_{v_i})$, $z_{\iota,i}^* = \Phi^{-1}(U_{z_{\iota,i}})(\iota = 1, \dots, p)$, $\Phi^{-1}(\cdot)$ is a quantile function of standard Gaussian, $\Phi_P(\cdot)$ is the joint CDF of a multivariate normal distribution with mean vector of zero and the covariance matrix Σ equal to the correlation matrix P. From (S.2) and (S.3), the joint density function of v_i and $z_i = (z_{1,i}, \dots, z_{p,i})'$ is

$$f(v_i, z_i') = \frac{1}{|P|^{\frac{1}{2}}} \exp\left[-\frac{(v_i^*, z_{1,i}^*, \dots, z_{p,i}^*)(P^{-1} - I_{p+1})(v_i^*, z_{1,i}^*, \dots, z_{p,i}^*)'}{2}\right] g(v_i) \prod_{\iota=1}^{p} h_\iota(z_{\iota,i}) \tag{S.4}$$

The corresponding log-likelihood function for a SAR model with endogenous W_n is

$$\ln L_n(\{v_i, z_i'\}|\lambda, \beta, \sigma_v^2, P) = -\frac{n}{2}\ln|P| - \frac{1}{2}\sum_{i=1}^{n}(v_i^*, z_{1,i}^*, \dots, z_{p,i}^*)(P^{-1} - I_{p+1})(v_i^*, z_{1,i}^*, \dots, z_{p,i}^*)' + \sum_{i=1}^{n} \ln \phi_{(0,\sigma_v^2)}(v_i(\lambda, \beta)) \tag{S.5}$$

where $\phi_{(0,\sigma_v^2)}(\cdot)$ is the normal density with mean 0 and variance σ_v^2. The nonparametric densities $h_\iota(z_{\iota,i})$ $(\iota = 1, \dots, p)$ disappear from the log-likelihood function as they do not include any parameters. For the endogenous heterogeneity specification, the scalar spatial coefficient λ should be replaced by the vector of parameters ζ. The pseudo-maximum likelihood estimation (PMLE) method in the following subsection is based on (S.5).

The Gaussian copula models $(v_i^*, z_{1,i}^*, \dots, z_{p,i}^*)'$ as the standard multivariate normal distribution with correlation matrix P. When there is one endogenous variable $z_{1,i}$, $(v_i^*, z_{1,i}^*)' \sim \mathrm{N}\left(\mathbf{0}_{2\times1}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$, which can be written as

$$\begin{pmatrix} v_i^* \\ z_{1,i}^* \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} \varpi_{1,i} \\ \varpi_{2,i} \end{pmatrix}$$

with $(\varpi_{1,i}, \varpi_{2,i})' \sim N(\mathbf{0}_{2\times1}, I_2)$, where I_2 is a 2×2 identity matrix. We have $v_i^* = \varpi_{1,i}$, $z_{1,i}^* = \rho\varpi_{1,i} + \sqrt{1-\rho^2}\varpi_{2,i}$, then $v_i^* = \frac{1}{\rho}z_{1,i}^* - \frac{\sqrt{1-\rho^2}}{\rho}\varpi_{2,i}$. Because $v_i = \Phi_{(0,\sigma_v^2)}^{-1}(\Phi(v_i^*)) = \sigma_v v_i^* = \gamma_1 z_{1,i}^* + \eta\varpi_{2,i}$, where $\Phi_{(0,\sigma_v^2)}(\cdot)$ is the normal distribution with mean 0 and variance σ_v^2, $\gamma_1 = \sigma_v \cdot \frac{1}{\rho}$, and $\eta = -\sigma_v \cdot \frac{\sqrt{1-\rho^2}}{\rho}$. A SAR model with endogeneous spatial weights (Eq.(1) in the main draft) can be rewritten as

$$y_i = \lambda \sum_{j\neq i} w_{ij}y_j + x_i'\beta + z_{1,i}^*\gamma_1 + \eta\varpi_{2,i}$$

where $z_{1,i}^*$ is an additional regressor to correct for the endogeneity bias, then W_n can be treated as predetermined or exogenous. The above approach can be easily extended to accommodate multiple endogenous variables $z_i^* = (z_{1,i}^*, ..., z_{p,i}^*)'$, i.e, $(v_i^*, z_i^{*'})' \sim N(\mathbf{0}_{(p+1)\times 1}, P)$. By Cholesky decomposition,

$$P = \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \cdot \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix}' = \begin{pmatrix} L_{11}L_{11}' & L_{11}L_{21}' \\ L_{21}L_{11}' & L_{21}L_{21}' + L_{22}L_{22}' \end{pmatrix}.$$

From this, we have that

$$\begin{pmatrix} v_i^* \\ z_i^* \end{pmatrix} = \begin{pmatrix} L_{11} = 1 & \mathbf{0}_{1\times p} \\ L_{21} = \rho_{vz} & L_{22} = \text{Chol}(P_z - \rho_{vz}\rho_{vz}') \end{pmatrix} \cdot \begin{pmatrix} \varpi_{1,i} \\ \varpi_{2,i} \end{pmatrix}$$

with $(\varpi_{1,i}, \varpi_{2,i}')' \sim N(\mathbf{0}_{(p+1)\times 1}, I_{p+1})$, where Chol$(\cdot)$ represents the Cholesky decomposition. Then we have the specification for the endogenous w_{ij} case below

$$\begin{aligned} y_i &= \lambda \sum_{j\neq i} w_{ij}y_j + x_i'\beta + \sigma_v v_i^* \\ &= \lambda \sum_{j\neq i} w_{ij}y_j + x_i'\beta + \sigma_v(\rho_{vz}'\rho_{vz})^{-1}\rho_{vz}' z_i^* - \sigma_v(\rho_{vz}'\rho_{vz})^{-1}\rho_{vz}'\text{Chol}(P_z - \rho_{vz}\rho_{vz}')\varpi_{2,i} \\ &= \lambda \sum_{j\neq i} w_{ij}y_j + x_i'\beta + z_i^{*\prime}\gamma + \epsilon_i \end{aligned} \tag{S.6}$$

where $\gamma = \sigma_v \rho_{vz}(\rho_{vz}'\rho_{vz})^{-1}$, $\epsilon_i = \eta\varpi_{p,i}$ with $\eta = -\sigma_v(\rho_{vz}'\rho_{vz})^{-1}\rho_{vz}'\text{Chol}(P_z - \rho_{vz}\rho_{vz}')$. Similarly, the specifications for Eq.(4) and (5) in the main draft are

$$y_i = \lambda(\zeta, z_i) \sum_{j\neq i} w_{ij}y_j + x_i'\beta + z_i^{*\prime}\gamma + \epsilon_i \tag{S.7}$$

and

$$y_i = \lambda \sum_{j\neq i} w_{ij}y_j + x_i'\beta_1 + z_i'\beta_2 + z_i^{*\prime}\gamma + \epsilon_i \tag{S.8}$$

respectively. (S.6)-(S.8) indicate that the models can also be estimated by an IV estimation approach to account for the endogenous W_nY_n.

S.1.2 Estimation

Denote $z_i^* = (z_{1,i}^*, \ldots, z_{p,i}^*)'$, by the partitioned quadratic formulation

$$(v_i^*, z_i^{*'})P^{-1}(v_i^*, z_i^{*'})' = (v_i^* - \rho_{vz}'P_z^{-1}z_i^*)'(1 - \rho_{vz}'P_z^{-1}\rho_{vz})^{-1}(v_i^* - \rho_{vz}'P_z^{-1}z_i^*) + z_i^{*\prime}P_z^{-1}z_i^*, \tag{S.9}$$

alternatively, the above log-likelihood function (S.5) can be written as

$$\begin{aligned} \ln L_n\left(\{v_i, z_i'\}\,|\lambda, \beta, \sigma_v^2, \kappa, \delta, P_z\right) = &-\frac{n}{2}\ln\kappa - \frac{n}{2}\ln|P_z| - \frac{1}{2}\sum_{i=1}^n \left[\frac{1}{\kappa}(v_i^* - z_i^*\delta)'(v_i^* - z_i^*\delta) + z_i^*(P_z^{-1} - I_p)z_i^* - v_i^{*\prime}v_i^*\right] \\ &+ \sum_{i=1}^n \ln\phi_{(0,\sigma_v^2)}\left(v_i(\lambda,\beta)\right) \end{aligned} \tag{S.10}$$

where $\kappa = 1 - \rho_{vz}' P_z^{-1} \rho_{vz}$ and $\delta = P_z^{-1} \rho_{vz}$. Since z_i^* in (S.6)-(S.8) and (S.10) are unobservables but can be consistently estimated, we propose a two-stage estimation method for the three model settings in the main draft.

S.1.2.1 The first stage estimation

In the first stage, we get estimates for the marginal transformations $\hbar_\iota(z) = \Phi^{-1}\left(H_\iota(z)\right)(\iota = 1, ..., p)$. For the purpose of this paper, any estimation method that yields an estimator $\hat{z}_{\iota,i}^*$ satisfying $\sup_{z_{\iota,i}} |\hat{z}_{\iota,i}^* - z_{\iota,i}^*| = o_p(1)$ can be chosen. Let $\hat{H}_\iota(z) = \frac{1}{n}\sum_{i=1}^n I(z_{\iota,i} \leq z)$ be the empirical distribution function of z_ι. As in the main draft, we consider the estimator proposed in Liu et al. (2012):

$$\hat{\hbar}_\iota(z) := \Phi^{-1}\left(T_{1/(2n)}[\hat{H}_\iota(z)]\right) \tag{S.11}$$

where $T_{1/(2n)}[\hat{H}_\iota(z)] := \frac{1}{2n} \cdot I(\hat{H}_\iota(z) < \frac{1}{2n}) + \hat{H}_\iota(z) \cdot I(\frac{1}{2n} \leq \hat{H}_\iota(z) \leq 1 - \frac{1}{2n}) + (1 - \frac{1}{2n}) \cdot I(\hat{H}_\iota(z) > 1 - \frac{1}{2n})$ is a Winsorization (or truncation) operator, the truncation level $\frac{1}{2n}$ is chosen to control the trade-off of bias and variance in high dimensions. Therefore, $\hat{z}_{\iota,i}^* = \hat{\hbar}_\iota(z_{\iota,i})$ for $i = 1, ..., n$ and $\iota = 1, ..., p$. By Theorem 4.6 in Liu et al. (2012), we have $\sup_{z_{\iota,i}} |\hat{z}_{\iota,i}^* - z_{\iota,i}^*| = o_p(1)$, the errors coming from the estimation of $z_{\iota,i}^*$ by $\hat{z}_{\iota,i}^*$ are asymptotically negligible.

S.1.2.2 The second stage estimation

S.1.2.3 2-Stage pseudo-maximum likelihood estimation

Denote $\omega = (\lambda, \beta')'$ (or $\omega = (\zeta', \beta')'$ for Eq.(4) in the main draft), given $\hat{z}_i^*$ from the first stage estimation, as $v_i^* = \Phi^{-1}\left(\Phi_{(0,\sigma_v^2)}(v_i(\omega))\right) = \frac{v_i(\omega)}{\sigma_v}$, note that

$$\begin{aligned}
&-\frac{1}{2}\sum_{i=1}^n \left[\frac{1}{\kappa}(\frac{v_i(\omega)}{\sigma_v} - \hat{z}_i^*\delta)'(\frac{v_i(\omega)}{\sigma_v} - \hat{z}_i^*\delta) + \hat{z}_i^*(P_z^{-1} - I_p)\hat{z}_i^* - \frac{v_i(\omega)^2}{\sigma_v^2}\right] \\
&= -\frac{1}{2\kappa\sigma_v^2}\sum_{i=1}^n (v_i(\omega) - \hat{z}_i^*\delta\sigma_v)'(v_i(\omega) - \hat{z}_i^*\delta\sigma_v) - \frac{1}{2}\sum_{i=1}^n \hat{z}_i^*(P_z^{-1} - I_p)\hat{z}_i^* + \frac{1}{2\sigma_v^2}\sum_{i=1}^n v_i(\omega)^2 \qquad \text{(S.12)} \\
&= -\frac{1}{2\kappa\sigma_v^2}(V_n(\omega) - \hat{Z}_n^*\delta\sigma_v)'(V_n(\omega) - \hat{Z}_n^*\delta\sigma_v) - \frac{1}{2}\sum_{i=1}^n \hat{z}_i^*(P_z^{-1} - I_p)\hat{z}_i^* + \frac{1}{2\sigma_v^2}V_n(\omega)'V_n(\omega),
\end{aligned}$$

and

$$\sum_{i=1}^n \ln \phi_{(0,\sigma_v^2)}(v_i(\omega)) = \ln \phi\left(\frac{Y_n - S_n^{-1}(\lambda)X_n\beta}{\sqrt{S_n^{-1}(\lambda)S_n^{-1\prime}(\lambda)\sigma_v^2}}\right) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_v^2 + \ln|S_n(\lambda)| - \frac{1}{2\sigma_v^2}V_n(\omega)'V_n(\omega),$$

where $V_n(\omega) = S_n(\lambda)Y_n - X_n\beta$. For the endogenous heterogeneity setting, $V_n(\omega) = S_n(\zeta)Y_n - X_n\beta$, where $S_n(\zeta) = I_n - \Lambda(\zeta, Z_n)W_n$ with $\Lambda(\zeta, Z_n) \equiv \text{diag}\{\lambda(\zeta, z_1), ..., \lambda(\zeta, z_n)\}$. Therefore, the log-likelihood function can be rewritten as

$$\ln L_n(\theta_{ML}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_\xi^2 - \frac{n}{2}\ln|P_z| + \ln|S_n(\lambda)| - \frac{1}{2\sigma_\xi^2}(V_n(\omega) - \hat{Z}_n^*\varphi)'(V_n(\omega) - \hat{Z}_n^*\varphi) - \frac{1}{2}\sum_{i=1}^n \hat{z}_i^{*\prime}(P_z^{-1} - I_p)\hat{z}_i^* \tag{S.13}$$

where $\theta_{ML} = (\omega', \sigma_\xi^2, \alpha', \varphi')'$ with $\sigma_\xi^2 = \kappa\sigma_v^2$, $\varphi = \delta\sigma_v$ and α is an $J(=\frac{(p-1)(p-2)}{2})$-dimensional column vector of distinct element in P_z. The 2SPML estimator $\hat{\theta}_{ML} = \arg\max \ln L_n(\theta_{ML})$.

For a SAR model with endogenous spatial weights, the first order derivatives at $\theta_{ML,0}^w$ are

$$\frac{\partial \ln L_n(\theta_{ML,0}^w)}{\partial\theta_{ML}} = \begin{pmatrix} \frac{\partial \ln L_n(\theta_{ML,0}^w)}{\partial\lambda} \\ \frac{\partial \ln L_n(\theta_{ML,0}^w)}{\partial\beta} \\ \frac{\partial \ln L_n(\theta_{ML,0}^w)}{\partial\sigma_\xi^2} \\ \frac{\partial \ln L_n(\theta_{ML,0}^w)}{\partial\alpha} \\ \frac{\partial \ln L_n(\theta_{ML,0}^w)}{\partial\varphi} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_{\xi,0}^2}(W_nY_n)'\left(V_n - \hat{Z}_n^*\varphi_0\right) - \text{tr}[W_nS_n^{-1}] \\ \frac{1}{\sigma_{\xi,0}^2}X_n'\left(V_n - \hat{Z}_n^*\varphi_0\right) \\ -\frac{n}{2\sigma_{\xi,0}^2} + \frac{1}{2\sigma_{\xi,0}^4}\left(V_n - \hat{Z}_n^*\varphi_0\right)'\left(V_n - \hat{Z}_n^*\varphi_0\right) \\ -\frac{n}{2}\frac{\partial\ln|P_{z,0}|}{\partial\alpha} - \frac{1}{2}\frac{\partial}{\partial\alpha}\text{tr}[P_{z,0}^{-1}\hat{Z}_n^{*\prime}\hat{Z}_n^*] \\ \frac{1}{\sigma_{\xi,0}^2}\hat{Z}_n^{*\prime}(V_n - \hat{Z}_n^*\varphi_0) \end{pmatrix}, \tag{S.14}$$

where the J-dimensional vector $\frac{\partial\ln|P_{z,0}|}{\partial\alpha}$ has the jth element $\text{tr}(P_{z,0}^{-1}\frac{\partial P_{z,0}}{\partial\alpha_j})$ and $\frac{\partial}{\partial\alpha}\text{tr}[P_{z,0}^{-1}\hat{Z}_n^{*\prime}\hat{Z}_n^*]$ has its jth element $-\text{tr}(P_{z,0}^{-1}\frac{\partial P_{z,0}}{\partial\alpha_j}P_{z,0}^{-1}\hat{Z}_n^{*\prime}\hat{Z}_n^*)$ for $j = 1, ..., J$. From items in (S.14) and the reduced form Eq.(3) in the main draft, denote $G_n = W_nS_n^{-1}$, we have $\frac{\partial \ln L_n(\theta_{ML,0}^w)}{\partial\theta_{ML}} = \frac{\partial \ln L_n^{u_w}(\theta_{ML,0})}{\partial\theta_{ML}} + \Lambda_n^w$, where

$$\frac{\partial \ln L_n^{u_w}(\theta_{ML,0})}{\partial\theta_{ML}} = \begin{pmatrix} \frac{1}{\sigma_{\xi,0}^2}[(G_nV_n)'(V_n - \hat{Z}_n^*\varphi_0)] - \text{tr}(G_n) \\ \frac{1}{\sigma_{\xi,0}^2}X_n'V_n \\ -\frac{n}{2\sigma_{\xi,0}^2} + \frac{1}{2\sigma_{\xi,0}^4}\left(V_n - \hat{Z}_n^*\varphi_0\right)'\left(V_n - \hat{Z}_n^*\varphi_0\right) \\ -\frac{n}{2}\frac{\partial\ln|P_{z,0}|}{\partial\alpha} - \frac{1}{2}\frac{\partial}{\partial\alpha}\text{tr}[P_{z,0}^{-1}\hat{Z}_n^{*\prime}\hat{Z}_n^*] \\ \frac{1}{\sigma_{\xi,0}^2}\hat{Z}_n^{*\prime}(V_n - \hat{Z}_n^*\varphi_0) \end{pmatrix}, \tag{S.15}$$

and

$$\Lambda_n^w = \left(-\frac{1}{\sigma_{\xi,0}^2}[(G_nX_n\beta_0)'(\hat{Z}_n^*\varphi_0)], -\frac{1}{\sigma_{\xi,0}^2}[X_n'(\hat{Z}_n^*\varphi_0)], 0, \mathbf{0}_{J\times1}', \mathbf{0}_{p\times1}'\right)'. \tag{S.16}$$

For a SAR model with endogenous heterogeneity, the differences are that $V_n(\omega) = S_n(\zeta)Y_n - X_n\beta$, where $S_n(\zeta) = I_n - \Lambda(\zeta, Z_n)W_n$ and ζ is a p_0-dimensional vector of parameters, and that $\frac{\partial\ln L_n(\theta_{ML,0})}{\partial\lambda}$ should be replaced by

$$\frac{\partial\ln L_n(\theta_{ML,0})}{\partial\zeta_\iota} = \frac{1}{\sigma_{\xi,0}^2}\left[\text{diag}\left(\frac{\partial\lambda(\zeta_0,z_i)}{\partial\zeta_\iota}\right)W_nY_n\right]'\left(V_n - \hat{Z}_n^*\varphi_0\right) - \text{tr}\left[\text{diag}\left(\frac{\partial\lambda(\zeta_0,z_i)}{\partial\zeta_\iota}\right)W_nS_n^{-1}\right], \ \iota = 1, ..., p_0.$$

By the reduced form, $Y_n = S_n^{-1}(X_n\beta_0 + V_n)$ with $S_n = I_n - \Lambda(\zeta_0, Z_n)W_n$, denote $\tilde{G}_{n,\iota} = \text{diag}(\frac{\partial\lambda(\zeta_0,z_i)}{\partial\zeta_\iota})G_n$

with $G_n = W_n S_n^{-1}$, we have $\frac{\partial \ln L_n(\theta^h_{ML,0})}{\partial \theta_{ML}} = \frac{\partial \ln L_n^{u_h}(\theta_{ML,0})}{\partial \theta_{ML}} + \Lambda_n^h$, where

$$\frac{\partial \ln L_n^{u_h}(\theta_{ML,0})}{\partial \theta_{ML}} = \begin{pmatrix} \frac{1}{\sigma^2_{\xi,0}}(\tilde{G}_{n,1}V_n)'(V_n - \hat{Z}_n^*\varphi_0) - \mathrm{tr}(\tilde{G}_{n,1}) \\ \vdots \\ \frac{1}{\sigma^2_{\xi,0}}(\tilde{G}_{n,p_0}V_n)'(V_n - \hat{Z}_n^*\varphi_0) - \mathrm{tr}(\tilde{G}_{n,p_0}) \\ \frac{1}{\sigma^2_{\xi,0}} X_n' V_n \\ -\frac{n}{2\sigma^2_{\xi,0}} + \frac{1}{2\sigma^4_{\xi,0}}\left(V_n - \hat{Z}_n^*\varphi_0\right)'\left(V_n - \hat{Z}_n^*\varphi_0\right) \\ -\frac{n}{2}\frac{\partial \ln|P_{z,0}|}{\partial \alpha} - \frac{1}{2}\frac{\partial}{\partial \alpha}\mathrm{tr}[P_{z,0}^{-1}\hat{Z}_n^{*\prime}\hat{Z}_n^*] \\ \frac{1}{\sigma^2_{\xi,0}}\hat{Z}_n^{*\prime}(V_n - \hat{Z}_n^*\varphi_0) \end{pmatrix}, \tag{S.17}$$

and

$$\Lambda_n^h = \left(-\frac{1}{\sigma^2_{\xi,0}}(\tilde{G}_{n,1}X_n\beta_0)'(\hat{Z}_n^*\varphi_0), \cdots, -\frac{1}{\sigma^2_{\xi,0}}(\tilde{G}_{n,p_0}X_n\beta_0)'(\hat{Z}_n^*\varphi_0), -\frac{1}{\sigma^2_{\xi,0}}X_n'(\hat{Z}_n^*\varphi_0), 0, \mathbf{0}'_{J\times 1}, \mathbf{0}'_{p\times 1}\right)'. \tag{S.18}$$

For a SAR model with endogenous regressors (Eq.(5) in the main draft), $V_n(\omega) = S_n(\lambda)Y_n - X_n\beta_1 - Z_n\beta_2$, the first order derivatives at $\theta^z_{ML,0}$ are

$$\frac{\partial \ln L_n(\theta^z_{ML,0})}{\partial \theta_{ML}} = \begin{pmatrix} \frac{\partial \ln L_n(\theta^z_{ML,0})}{\partial \lambda} \\ \frac{\partial \ln L_n(\theta^z_{ML,0})}{\partial \beta_1} \\ \frac{\partial \ln L_n(\theta^z_{ML,0})}{\partial \beta_2} \\ \frac{\partial \ln L_n(\theta^z_{ML,0})}{\partial \sigma^2_\xi} \\ \frac{\partial \ln L_n(\theta^z_{ML,0})}{\partial \alpha} \\ \frac{\partial \ln L_n(\theta^z_{ML,0})}{\partial \varphi} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2_{\xi,0}}(W_nY_n)'\left(V_n - \hat{Z}_n^*\varphi_0\right) - \mathrm{tr}[W_nS_n^{-1}] \\ \frac{1}{\sigma^2_{\xi,0}}X_n'\left(V_n - \hat{Z}_n^*\varphi_0\right) \\ \frac{1}{\sigma^2_{\xi,0}}Z_n'\left(V_n - \hat{Z}_n^*\varphi_0\right) \\ -\frac{n}{2\sigma^2_{\xi,0}} + \frac{1}{2\sigma^4_{\xi,0}}\left(V_n - \hat{Z}_n^*\varphi_0\right)'\left(V_n - \hat{Z}_n^*\varphi_0\right) \\ -\frac{n}{2}\frac{\partial \ln|P_{z,0}|}{\partial \alpha} - \frac{1}{2}\frac{\partial}{\partial \alpha}\mathrm{tr}[P_{z,0}^{-1}\hat{Z}_n^{*\prime}\hat{Z}_n^*] \\ \frac{1}{\sigma^2_{\xi,0}}\hat{Z}_n^{*\prime}(V_n - \hat{Z}_n^*\varphi_0) \end{pmatrix}, \tag{S.19}$$

then we have $\frac{\partial \ln L_n(\theta^z_{ML,0})}{\partial \theta_{ML}} = \frac{\partial \ln L_n^{u_z}(\theta_{ML,0})}{\partial \theta_{ML}} + \Lambda_n^z$, where

$$\frac{\partial \ln L_n^{u_z}(\theta_{ML,0})}{\partial \theta_{ML}} = \begin{pmatrix} \frac{1}{\sigma^2_{\xi,0}}[(G_nV_n)'(V_n - \hat{Z}_n^*\varphi_0)] - \mathrm{tr}(G_n) \\ \frac{1}{\sigma^2_{\xi,0}}X_n'V_n \\ \mathbf{0}_{p\times 1} \\ -\frac{n}{2\sigma^2_{\xi,0}} + \frac{1}{2\sigma^4_{\xi,0}}\left(V_n - \hat{Z}_n^*\varphi_0\right)'\left(V_n - \hat{Z}_n^*\varphi_0\right) \\ -\frac{n}{2}\frac{\partial \ln|P_{z,0}|}{\partial \alpha} - \frac{1}{2}\frac{\partial}{\partial \alpha}\mathrm{tr}[P_{z,0}^{-1}\hat{Z}_n^{*\prime}\hat{Z}_n^*] \\ \frac{1}{\sigma^2_{\xi,0}}\hat{Z}_n^{*\prime}(V_n - \hat{Z}_n^*\varphi_0) \end{pmatrix}, \tag{S.20}$$

and

$$\Lambda_n^z = \left(-\frac{1}{\sigma^2_{\xi,0}}\left[(G_n(X_n\beta_{1,0} + Z_n\beta_{2,0}))'\left(\hat{Z}_n^*\varphi_0\right)\right], -\frac{1}{\sigma^2_{\xi,0}}[X_n'(\hat{Z}_n^*\varphi_0)], -\frac{1}{\sigma^2_{\xi,0}}[Z_n'(V_n - Z_n^*\varphi_0)], 0, \mathbf{0}'_{J\times 1}, \mathbf{0}'_{p\times 1}\right)'. \tag{S.21}$$

Remark 1 (the impacts of the two assumptions for 2SPMLE). First, we consider the first and second variants of a SAR model. Assumption S.2 are not required for the identification of those two settings because the endogenous z_i are not included as explanatory variables directly. Although $\mathrm{E}\left(\frac{\partial \ln L_n^{uw}(\theta_{ML,0})}{\partial \theta_{ML}}\right) = \mathbf{0}_{(k+p+J+2)\times 1}$ and $\mathrm{E}\left(\frac{\partial \ln L_n^{uh}(\theta_{ML,0})}{\partial \theta_{ML}}\right) = \mathbf{0}_{(k+2p+J+1)\times 1}$, Assumption S.1 determines whether the two parts $\mathrm{E}(\Lambda_n^w)$ and $\mathrm{E}(\Lambda_n^h)$ would cause asymptotic bias of the 2SPML estimator or not. When Assumption S.1 holds, x_i and z_i^* are uncorrelated, $\mathrm{E}[(G_nX_n\beta_0)'(\hat{Z}_n^*\varphi_0)] = 0$, $\mathrm{E}[X_n'(\hat{Z}_n^*\varphi_0)] = \mathbf{0}_{k\times 1}$, and $\mathrm{E}[(\tilde{G}_{n,1}X_n\beta_0)'(\hat{Z}_n^*\varphi_0)] = 0$, $\cdots$, $\mathrm{E}[(\tilde{G}_{n,p}X_n\beta_0)'(\hat{Z}_n^*\varphi_0)] = 0$, which implies $\mathrm{E}(\Lambda_n^w) = \mathbf{0}_{(k+p+J+2)\times 1}$ and $\mathrm{E}(\Lambda_n^h) = \mathbf{0}_{(k+2p+J+1)\times 1}$, then $\frac{\partial \ln L_n(\theta_{ML,0}^w)}{\partial \theta_{ML}}$ and $\frac{\partial \ln L_n(\theta_{ML,0}^h)}{\partial \theta_{ML}}$ have zero mean and are asymptotically normally distributed. Otherwise, when Assumption S.1 is violated, i.e., x_i and z_i^* are correlated, $\mathrm{E}[(G_nX_n\beta_0)'(\hat{Z}_n^*\varphi_0)] \neq 0$, $\mathrm{E}[X_n'(\hat{Z}_n^*\varphi_0)] \neq \mathbf{0}_{k\times 1}$, and $\mathrm{E}[(\tilde{G}_{n,1}X_n\beta_0)'(\hat{Z}_n^*\varphi_0)] \neq 0$, $\cdots$, $\mathrm{E}[(\tilde{G}_{n,p}X_n\beta_0)'(\hat{Z}_n^*\varphi_0)] \neq 0$, then $\mathrm{E}(\Lambda_n^w) \neq \mathbf{0}_{(k+p+J+2)\times 1}$ and $\mathrm{E}(\Lambda_n^h) \neq \mathbf{0}_{(k+2p+J+1)\times 1}$, which will cause the asymptotic bias of the 2SMPL estimator. Second, we consider the third variant of a SAR model. We need both Assumption S.1 and Assumption S.2 to guarantee the consistency of the 2SMPL estimator. To see this, if Assumption S.1 doesn't hold, although $\mathrm{E}\left(\frac{\partial \ln L_n^{uz}(\theta_{ML,0})}{\partial \theta_{ML}}\right) = \mathbf{0}_{(k+2p+J+2)\times 1}$, the part Λ_n^z might cause asymptotic bias as $\mathrm{E}[(G_nX_n\beta_{1,0})'(\hat{Z}_n^*\varphi_0)] \neq 0$, $\mathrm{E}[X_n'(\hat{Z}_n^*\varphi_0)] \neq \mathbf{0}_{k\times 1}$ (and also because $\mathrm{E}[(G_nZ_n\beta_{2,0})'(\hat{Z}_n^*\varphi_0)] \neq 0$ and it's possible that $\mathrm{E}[Z_n'(V_n - Z_n^*\varphi_0)] \neq \mathbf{0}_{p\times 1}$). When Assumption S.2 doesn't hold, i.e., for $\iota = 1, ..., p$, each $z_{\iota,i}$ follows a normal distribution, $z_{\iota,i}^*$ is a linear transformation of $z_{\iota,i}$ because $z_{\iota,i}^* = \Phi^{-1}(H_\iota(z_{\iota,i}))$. Then $V_n(\omega) - \hat{Z}_n^*\varphi$ in the log-likelihood function (S.13) becomes $S_n(\lambda)Y_n - X_n\beta_1 - Z_n\beta_2 - \hat{Z}_n^*\varphi$ and $\frac{\partial \ln L_n(\theta_{ML,0}^z)}{\partial \beta_2}$ and $\frac{\partial \ln L_n(\theta_{ML,0}^z)}{\partial \varphi}$ might be proportional to each other, we can't separately identify β_2 and φ.

S.1.2.4 2-Stage IV estimation

Given $\hat{Z}_n^*$ from the first stage estimation, by substituting $\hat{Z}_n^*$ for Z_n^* in (S.6)-(S.8), we have

$$Y_n = \lambda W Y_n + X_n\beta + \hat{Z}_n^*\gamma + \hat{\boldsymbol{\epsilon}}_n, \tag{S.22}$$

$$Y_n = \Lambda(\zeta, Z_n) W Y_n + X_n\beta + \hat{Z}_n^*\gamma + \hat{\boldsymbol{\epsilon}}_n, \tag{S.23}$$

$$Y_n = \lambda W Y_n + X_n\beta_1 + Z_n\beta_2 + \hat{Z}_n^*\gamma + \hat{\boldsymbol{\epsilon}}_n, \tag{S.24}$$

where $\hat{\boldsymbol{\epsilon}}_n = \boldsymbol{\epsilon}_n + (Z_n^* - \hat{Z}_n^*)\gamma$. We consider the IV estimators for the three models separately. For a SAR model with endogenous W_n (S.22), denote $\theta_{IV,w} = (\lambda, \beta', \gamma')'$, $\hat{M}_{n,w} = \left(W_nY_n, X_n, \hat{Z}_n^*\right)$ and $\hat{T}_{n,w} = \left(Q_{n,w}, X_n, \hat{Z}_n^*\right)$, where $Q_{n,w}$ is an instrument variable matrix for the endogenous W_nY_n, for example, the column vectors of $Q_{n,w}$ can be linear combinations of $X_n, W_nX_n, W_n^2X_n, ...$ and columns in $\hat{Z}_n^*$. The 2SIV estimator of $\theta_{IV,w}$ is

$$\hat{\theta}_{IV,w} = \left[\hat{M}_{n,w}'\hat{T}_{n,w}\left(\hat{T}_{n,w}'\hat{T}_{n,w}\right)^{-1}\hat{T}_{n,w}'\hat{M}_{n,w}\right]^{-1}\hat{M}_{n,w}'\hat{T}_{n,w}\left(\hat{T}_{n,w}'\hat{T}_{n,w}\right)^{-1}\hat{T}_{n,w}'Y_n \tag{S.25}$$

For a SAR model with endogenous heterogeneity (S.23), the IV estimation approach can only be applied to the case when $\lambda(\zeta, z_i) = \varrho_1 F_1(z_{1,i}) + \cdots + \varrho_p F_p(z_{p,i})$, where $F_\iota(\cdot)$ $(\iota = 1, ..., p)$ are some globally bounded functions, e.g., continuous probability functions. Denote $\theta_{IV,h} = (\zeta', \beta', \gamma')'$ with $\zeta = (\varrho_1, ..., \varrho_p)'$, $\hat{M}_{n,h} = \left(\Lambda_1(z_1)W_nY_n, ..., \Lambda_p(z_p)W_nY_n, X_n, \hat{Z}_n^*\right)$ with $\Lambda_\iota(z_\iota) = \text{diag}\{F_\iota(z_{\iota,1}), ..., F_\iota(z_{\iota,n})\}(\iota = 1, ..., p)$, $\hat{T}_{n,h} = (Q_{1n,h}, ..., Q_{pn,h}, X_n, \hat{Z}_n^*)$, where $Q_{\iota n,h}$ is an instrument variable matrix for $\Lambda_\iota(z_\iota)W_nY_n$, for instance, instruments $Q_{1n,h}, ..., Q_{pn,h}$ may be constructed as subsets of the linearly independent columns of

$$X_n, \Lambda_1(z_1)W_nX_n, (\Lambda_1(z_1)W_n)^2X_n, ..., \Lambda_2(z_2)W_nX_n, (\Lambda_2(z_2)W_n)^2X_n, ..., \Lambda_p(z_p)W_nX_n, (\Lambda_p(z_p)W_n)^2X_n, ...$$

and columns in $\hat{Z}_n^*$. [1] Denote $Q_{n,h} = (Q_{1n,h}, ..., Q_{pn,h})$. The 2SIV estimator of $\theta_{IV,h}$ is

$$\hat{\theta}_{IV,h} = \left[\hat{M}'_{n,h}\hat{T}_{n,h}\left(\hat{T}'_{n,h}\hat{T}_{n,h}\right)^{-1}\hat{T}'_{n,h}\hat{M}_{n,h}\right]^{-1}\hat{M}'_{n,h}\hat{T}_{n,h}\left(\hat{T}'_{n,h}\hat{T}_{n,h}\right)^{-1}\hat{T}'_{n,h}Y_n. \tag{S.26}$$

For a SAR model with endogenous regressor Z_n (S.24), denote $\theta_{IV,z} = (\lambda, \beta_1', \beta_2', \gamma')'$, $\hat{M}_{n,z} = \left(W_nY_n, X_n, Z_n, \hat{Z}_n^*\right)$ and $\hat{T}_{n,z} = \left(Q_n, X_n, Z_n, \hat{Z}_n^*\right)$, where $Q_{n,z}$ is an instrument variable matrix for the endogenous W_nY_n, for example, $Q_{n,z}$ can be linear combinations of $X_n, W_nX_n, W_n^2X_n, ...$ and columns in Z_n and $\hat{Z}_n^*$. The 2SIV estimator of $\theta_{IV,z}$ is

$$\hat{\theta}_{IV,z} = \left[\hat{M}'_{n,z}\hat{T}_{n,z}\left(\hat{T}'_{n,z}\hat{T}_{n,z}\right)^{-1}\hat{T}'_{n,z}\hat{M}_{n,z}\right]^{-1}\hat{M}'_{n,z}\hat{T}_{n,z}\left(\hat{T}'_{n,z}\hat{T}_{n,z}\right)^{-1}\hat{T}'_{n,z}Y_n. \tag{S.27}$$

Remark 2 (the impacts of the two assumptions for 2SIVE). Denote $G_n = W_nS_n^{-1}$ and $\tilde{G}_{n,\iota} = \Lambda_\iota(z_\iota)G_n$ for $\iota = 1, ..., p$. For identification of the 2SIV estimators, two assumptions are needed (Kelejian and Prucha, 1998): $\lim_{n\to\infty}\frac{1}{n}\text{E}(Q'_{n,\cdot}Q_{n,\cdot})$ is nonsingular and $\lim_{n\to\infty}\frac{1}{n}\text{E}(Q'_{n,\cdot}\hat{\mathcal{M}}_{n,\cdot})$ has full column rank, where $\hat{\mathcal{M}}_{n,w} = (G_n(X_n\beta_0 + Z_n^*\gamma_0), X_n, Z_n^*)$, $\hat{\mathcal{M}}_{n,h} = \left(\tilde{G}_{n,1}(X_n\beta_0 + Z_n^*\gamma_0), ..., \tilde{G}_{n,p}(X_n\beta_0 + Z_n^*\gamma_0), X_n, Z_n^*\right)$, and $\hat{\mathcal{M}}_{n,z} = (G_n(X_n\beta_0 + Z_n^*\gamma_0), X_n, Z_n, Z_n^*)$. When Assumption S.1 holds for (S.22)-(S.23), and Assumptions S.1-S.2 hold for (S.24), the two identification assumptions would be satisfied. Otherwise, if Assumptions S.1-S.2 are violated, i.e., X_n and Z_n^* are correlated and/or Z_n is normally distributed so that Z_n^* are linear transformations of Z_n, then $\lim_{n\to\infty}\frac{1}{n}\text{E}(Q'_{n,\cdot}Q_{n,\cdot})$ can be singular and $\lim_{n\to\infty}\frac{1}{n}\text{E}(Q'_{n,\cdot}\hat{\mathcal{M}}_{n,\cdot})$ would not have full column rank.

S.2 A sufficient identification result for a SAR model with endogenous W_n

Denote $\boldsymbol{X}_n = \left(X_n, \left(\mathscr{O}_n^{\perp}Z_n^*\right)\right)$, for MLE with a finite sample, identification is equivalent to $P(\ln L_n(\theta_{ML,0}) \neq \ln L_n(\theta_{ML})) > 0$ for any $\theta_{ML,0} \neq \theta_{ML}$ (Rothenberg, 1971). Below is a sufficient identification result.

Lemma S.1. *Under Assumptions 1, 2 and 4 in the main draft, if $\boldsymbol{X}_n'\boldsymbol{X}_n$ is invertible with probability 1, $W_n + W_n' \neq 0$, and there exists $j \neq j'$ such that $\sum_{i=1}^n w_{ij}^2 \neq \sum_{i=1}^n w_{ij'}^2$, then $\theta_{ML,0} = (\lambda_0, \beta_0', \sigma_\xi^2, \chi_0', \alpha_0', \delta_0')'$ is identified.*

1. We may also employ columns of X_n pre-multiplied by cross-products of the $\Lambda_\iota(z_\iota)W_n$ and columns in $\hat{Z}_n^*$.

Proof. By the similar argument in the proof of Lemma 3 in the main draft, we have $\frac{1}{n}\ln L_{n0}(\theta_{ML}) + o_p(1) = \frac{1}{n}\ln L_{n0}(\theta_{ML,0}) + o_p(1)$ almost surely, i.e.,

$$\begin{aligned}
&-\frac{n}{2}\ln\sigma_\xi^2 - \frac{n}{2}\ln|P_x| - \frac{n}{2}\ln|\Xi| + \ln|I_n - \lambda W_n| - \frac{1}{2}\sum_{i=1}^n (z_i^* - \Gamma' x_i^*)' \,\Xi^{-1}(z_i^* - \Gamma' x_i^*) - \frac{1}{2}\sum_{i=1}^n x_i^{*\prime} P_x^{-1} x_i^* \\
&- \frac{1}{2\sigma_\xi^2}[(I_n - \lambda W_n)Y_n - X_n\beta - \left(\mathscr{O}_n^\perp Z_n^*\right)\chi]'[(I_n - \lambda W_n)Y_n - X_n\beta - \left(\mathscr{O}_n^\perp Z_n^*\right)\chi] \\
=& -\frac{n}{2}\ln\sigma_{\xi,0}^2 - \frac{n}{2}\ln|P_{x,0}| - \frac{n}{2}\ln|\Xi_0| + \ln|I_n - \lambda_0 W_n| - \frac{1}{2}\sum_{i=1}^n (z_i^* - \Gamma_0' x_i^*)' \,\Xi_0^{-1}(z_i^* - \Gamma_0' x_i^*) - \frac{1}{2}\sum_{i=1}^n x_i^{*\prime} P_{x,0}^{-1} x_i^* \\
&- \frac{1}{2\sigma_{\xi,0}^2}[(I_n - \lambda_0 W_n)Y_n - X_n\beta_0 - \left(\mathscr{O}_n^\perp Z_n^*\right)\chi_0]'[(I_n - \lambda_0 W_n)Y_n - X_n\beta_0 - \left(\mathscr{O}_n^\perp Z_n^*\right)\chi_0]
\end{aligned} \tag{S.28}$$

holds for Y_n, Z_n, and X_n almost surely. Differentiate Eq.(S.28) with respect to y_j, we have

$$\begin{aligned}
&\sigma_\xi^{-2}\left\{y_j - \lambda w_{j\cdot}Y_n - x_j\beta - (\mathscr{O}_n^\perp Z_n^*)_j\chi - \lambda\sum_{i=1}^n\left[y_i - \lambda w_{i\cdot}Y_n - x_i\beta - \left(\mathscr{O}_n^\perp Z_n^*\right)_i\chi\right]w_{ij}\right\} \\
&= \sigma_{\xi,0}^{-2}\left\{y_j - \lambda_0 w_{j\cdot}Y_n - x_j\beta_0 - (\mathscr{O}_n^\perp Z_n^*)_j\chi_0 - \lambda_0\sum_{i=1}^n\left[y_i - \lambda_0 w_{i\cdot}Y_n - x_i\beta_0 - \left(\mathscr{O}_n^\perp Z_n^*\right)_i\chi_0\right]w_{ij}\right\}.
\end{aligned} \tag{S.29}$$

Differentiate Eq.(S.29) with respect to y_j once more,

$$\sigma_\xi^{-2}(1 - \lambda w_{jj} + \lambda^2\sum_{i=1}^n w_{ij}^2) = \sigma_{\xi,0}^{-2}(1 - \lambda_0 w_{jj} + \lambda_0^2\sum_{i=1}^n w_{ij}^2),$$

since $w_{jj} = 0$ and there exists $j \neq j'$ such that $\sum_{i=1}^n w_{ij}^2 \neq \sum_{i=1}^n w_{ij'}^2$, we have that $1/\sigma_\xi^2 = 1/\sigma_{\xi,0}^2$ and $\lambda^2/\sigma_\xi^2 = \lambda_0^2/\sigma_{\xi,0}^2$. Hence, $\sigma_\xi = \sigma_{\xi,0}$ and $|\lambda| = |\lambda_0|$. Differentiate Eq.(S.29) with respect to $y_k (k \neq j)$,

$$\sigma_\xi^{-2}(\lambda^2\sum_{i=1}^n w_{ik}w_{ij} - \lambda(w_{kj} + w_{jk})) = \sigma_{\xi,0}^{-2}(\lambda_0^2\sum_{i=1}^n w_{ik}w_{ij} - \lambda_0(w_{kj} + w_{jk})),$$

Thus, $\lambda(w_{kj} + w_{jk}) = \lambda_0(w_{kj} + w_{jk})$. Because $W_n + W_n' \neq 0$ and $w_{ii} = 0$, we have $\lambda = \lambda_0$. Eq.(S.29) implies that

$$\begin{aligned}
&(x_j\beta - \lambda\sum_{i=1}^n w_{ij}x_i\beta) + [(\mathscr{O}_n^\perp Z_n^*)_j\chi - \lambda\sum_{i=1} w_{ij}\left(\mathscr{O}_n^\perp Z_n^*\right)_i\chi] \\
&\quad = (x_j\beta_0 - \lambda_0\sum_{i=1}^n w_{ij}x_i\beta_0) + [(\mathscr{O}_n^\perp Z_n^*)_j\chi_0 - \lambda_0\sum_{i=1} w_{ij}\left(\mathscr{O}_n^\perp Z_n^*\right)_i\chi_0],
\end{aligned}$$

which is equivalent to $(I_n - \lambda_0 W_n')X_n\beta + (I_n - \lambda_0 W_n')(\mathscr{O}_n^\perp Z_n^*)\chi = (I_n - \lambda_0 W_n')X_n\beta_0 + (I_n - \lambda_0 W_n')(\mathscr{O}_n^\perp Z_n^*)\chi_0$. As $(I_n - \lambda_0 W_n')$ is invertible, X_n and $\mathscr{O}_n^\perp Z_n^*$ are not linearly independent, it must be that $X_n\beta = X_n\beta_0$ and $(\mathscr{O}_n^\perp Z_n^*)\chi = (\mathscr{O}_n^\perp Z_n^*)\chi_0$. Therefore, $\beta = \beta_0$ and $\chi = \chi_0$. The identification of Ξ and P_x (or the related distinct elements α and δ) follows the similar argument for the proof (Lemma 3) in the main draft for the endogenous heterogeneity case. □

S.3 Taylor expansions of $\ln|I_n - \lambda W_n|$ and $\ln|I_n - \Lambda(\zeta, Z_n)W_n|$

First, $\ln|I_n - \lambda W_n| = -\sum_{l=1}^{\infty}\frac{\lambda^l}{l}\text{tr}(W_n^l) = -\sum_{l=1}^{\infty}\frac{\lambda^l}{l}\sum_{i=1}^{n}(W_n^l)_{ii}$, the proof of which is provided in Appendix A of Qu and Lee (2013). Second, we consider the Taylor series of $\mathscr{P}(\zeta) = \ln|I_n - \Lambda(\zeta, Z_n)W_n|$. As $\Lambda(\zeta, Z_n) \equiv \text{diag}\{\lambda(\zeta, z_1), ..., \lambda(\zeta, z_n)\} = \text{diag}(\lambda(\zeta, z_i))$, under the assumptions that $\lambda(\zeta, \cdot)$ is smooth and strict monotonic for all elements in ζ and that $\lambda(\zeta^0, Z_n) \equiv 0$ with $\zeta^0 = \mathbf{0}_{p_0\times 1}$, we can use the Taylor expansion $\mathscr{P}(\zeta) = \mathscr{P}(\zeta^0) + \sum_{l=1}^{\infty}\sum_{l_1+\cdots+l_{p_0}=l}\frac{\zeta_1^{l_1}\cdots\zeta_{p_0}^{l_{p_0}}}{l_1!\cdots l_{p_0}!}\cdot\frac{\partial^{l_1}\cdots\partial^{l_{p_0}}\mathscr{P}(\zeta^0)}{\partial\zeta_1^{l_1}\cdots\partial\zeta_{p_0}^{l_{p_0}}}$ where $0 \le l_1, ..., l_{p_0} \le l$. As $\mathscr{P}(\zeta^0) = \ln|I_n| = 0$, it remains to consider the derivatives. Notice that

$$\frac{\partial\mathscr{P}(\zeta)}{\partial\zeta_\iota} = -\text{tr}\left[\text{diag}\left(\frac{\partial\lambda(\zeta, z_i)}{\partial\zeta_\iota}\right)W_n S_n^{-1}(\zeta)\right] = -\text{tr}\left[\text{diag}\left(\frac{\partial\lambda(\zeta, z_i)}{\partial\lambda_p}\right)W_n\sum_{j=0}^{\infty}(\Lambda(\zeta, Z_n)W_n)^j\right]$$
$$= -\text{tr}\left[\text{diag}(\frac{\partial\lambda(\zeta, z_i)}{\partial\zeta_\iota})W_n\sum_{j=1}^{\infty}(\Lambda(\zeta, Z_n)W_n)^j\right],\ \iota = 1, .., p_0,$$

the last equality holds because $\text{tr}\left[\text{diag}\left(\frac{\partial\lambda(\zeta,z_i)}{\partial\zeta_\iota}\right)W_n\right] = 0$. For any $\iota_1, \iota_2 = 1, ..., p_0$, we have

$$\frac{\partial^2\mathscr{P}(\zeta)}{\partial\zeta_{\iota_1}\partial\zeta_{\iota_2}} = -\text{tr}\left[\text{diag}\left(\frac{\partial^2\lambda(\zeta, z_i)}{\partial\zeta_{\iota_1}\partial\zeta_{\iota_2}}\right)W_n\sum_{j=1}^{\infty}(\Lambda(\zeta, Z_n)W_n)^j\right]$$
$$-\text{tr}\left[\text{diag}\left(\frac{\partial\lambda(\zeta, z_i)}{\partial\zeta_{\iota_1}}\cdot\frac{\partial\lambda(\zeta, z_i)}{\partial\zeta_{\iota_2}}\right)W_n^2\sum_{j=1}^{\infty}j(\Lambda(\zeta, Z_n)W_n)^{j-1}\right],\ \iota_1, \iota_2 = 1, ..., p_0,$$

and

$$\frac{\partial^3\mathscr{P}(\zeta)}{\partial\zeta_{\iota_1}\partial\zeta_{\iota_2}\partial\zeta_{\iota_3}} = -\text{tr}\left[\text{diag}\left(\frac{\partial^3\lambda(\zeta, z_i)}{\partial\zeta_{\iota_1}\partial\zeta_{\iota_2}\partial\zeta_{\iota_3}}\right)W_n\sum_{j=1}^{\infty}(\Lambda(\zeta, Z_n)W_n)^j\right]$$
$$-\text{tr}\left[\text{diag}\left(\frac{\partial^2\lambda(\zeta, z_i)}{\partial\zeta_{\iota_1}\partial\zeta_{\iota_2}}\cdot\frac{\partial\lambda(\zeta, z_i)}{\partial\zeta_{\iota_3}}\right)W_n^2\sum_{j=1}^{\infty}j(\Lambda(\zeta, Z_n)W_n)^{j-1}\right]$$
$$-\text{tr}\left[\text{diag}\left(\frac{\partial^2\lambda(\zeta, z_i)}{\partial\zeta_{\iota_1}\partial\zeta_{\iota_3}}\cdot\frac{\partial\lambda(\zeta, z_i)}{\partial\iota_{\iota_2}} + \frac{\partial\lambda(\zeta, z_i)}{\partial\zeta_{\iota_1}}\cdot\frac{\partial^2\lambda(\zeta, z_i)}{\partial\zeta_{\iota_2}\partial\zeta_{\iota_3}}\right)W_n^2\sum_{j=1}^{\infty}j(\Lambda(\zeta, Z_n)W_n)^{j-1}\right]$$
$$-\text{tr}\left[\text{diag}\left(\frac{\partial\lambda(\zeta, z_i)}{\partial\zeta_{\iota_1}}\cdot\frac{\partial\lambda(\zeta, z_i)}{\partial\zeta_{\iota_2}}\cdot\frac{\partial\lambda(\zeta, z_i)}{\partial\zeta_{\iota_3}}\right)W_n^3\sum_{j=2}^{\infty}j(j-1)(\Lambda(\zeta, Z_n)W_n)^{j-2}\right],\ \iota_1, \iota_2, \iota_3 = 1, ..., p_0.$$

At ζ^0,

$$\frac{\partial\mathscr{P}(\zeta^0)}{\partial\zeta_\iota} = 0,\ \frac{\partial^2\mathscr{P}(\zeta^0)}{\partial\zeta_{\iota_1}\partial\zeta_{\iota_2}} = -\sum_{i=1}^{n}\left[\frac{\partial\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_1}}\frac{\partial\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_2}}\left(W_n^2\right)_{ii}\right],$$
$$\frac{\partial^3\mathscr{P}(\zeta^0)}{\partial\zeta_{\iota_1}\partial\zeta_{\iota_2}\partial\zeta_{\iota_3}} = -\sum_{i=1}^{n}\left[\left(\frac{\partial^2\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_1}\partial\zeta_{\iota_2}}\frac{\partial\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_3}} + \frac{\partial^2\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_1}\partial\zeta_{\iota_3}}\frac{\partial\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_2}} + \frac{\partial\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_1}}\frac{\partial^2\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_2}\partial\zeta_{\iota_3}}\right)(W_n^2)_{ii}\right.$$
$$\left.-\sum_{i=1}^{n}\left[\frac{\partial\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_1}}\frac{\partial\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_2}}\frac{\partial\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_3}}(W_n^3)_{ii}\right].\right.$$

By induction, $\frac{\partial^{l_1}\cdots\partial^{l_{p_0}}\mathscr{P}(\zeta^0)}{\partial\zeta_1^{l_1}\cdots\partial\zeta_{p_0}^{l_{p_0}}} = -\sum_{i=1}^n \left[\mathcal{L}_i^{(2)}(\zeta^0, z_i, l_1, ..., l_{p_0})(W_n^2)_{ii} + \cdots + \mathcal{L}_i^{(l)}(\zeta^0, z_i, l_1, ..., l_{p_0})(W_n^l)_{ii}\right]$, where $\mathcal{L}_i^{(h)}(\zeta^0, z_i, l_1, ..., l_{p_0}), h = 1, ..., l$ are some combinations of the first, second, and higher order $(< l)$ partial derivatives of $\lambda(\zeta, \cdot)$ evaluated at ζ^0 and obviously

$$\mathcal{L}_i^{(l)}(\zeta^0, z_i, l_1, ..., l_{p_0}) = \underbrace{\frac{\partial\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_1}}\cdots\frac{\partial\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_1}}}_{l_1}\cdots\underbrace{\frac{\partial\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_{p_0}}}\cdots\frac{\partial\lambda(\zeta^0, z_i)}{\partial\zeta_{\iota_{p_0}}}}_{l_{p_0}}.$$

Therefore, $\ln|I_n - \Lambda(\zeta, Z_n)W_n| = -\sum_{l=1}^{\infty}\sum_{l_1+\cdots+l_{p_0}=l}\frac{\zeta_1^{l_1}\cdots\zeta_{p_0}^{l_{p_0}}}{l_1!\cdots l_{p_0}!}\sum_{i=1}^n\left[\mathcal{L}_i^{(2)}(\zeta^0, z_i, l_1, ..., l_{p_0})(W_n^2)_{ii} + \cdots\right.$ $\left.+\mathcal{L}_i^{(l)}(\zeta^0, z_i, l_1, ..., l_{p_0})(W_n^l)_{ii}\right]$.

S.4 Useful properties for a SAR model with endogenous heterogeneity

The following lemma is about the uniform convergence of the log-determinant term in the log pseudo-likelihood function.

Lemma S.2. *Under Assumptions 4(i)-(iii) and 5 in the main draft,* $\sup_{\zeta\in\Theta_\zeta}\frac{1}{n}(\ln|I_n - \Lambda(\zeta, Z_n)W_n| - \mathrm{E}\ln|I_n - \Lambda(\zeta, Z_n)W_n|) \xrightarrow{p} 0$.

Proof. First, we consider the pointwise convergence $\mathrm{plim}_{n\to\infty}\frac{1}{n}(\ln|I_n - \Lambda(\zeta, Z_n)W_n| - \mathrm{E}\ln|I_n - \Lambda(\zeta, Z_n)W_n|) = 0$, by the WLLN in Jenish and Prucha (2012), we only need to check the NED property and uniform L_p boundedness of $\ln|I_n - \Lambda(\zeta, Z_n)W_n|$. Based on the Taylor expansion in Section S.3, we show that $\mathcal{L}_i^{(h)}(\zeta^0, z_i, l_1, \ldots, l_{p_0})(W_n^h)_{ii} \leq \tilde{c}_i^{*(h)}s^{-c_0d_0}$ 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 $\{w_{ij_1}, w_{j_1j_2}, ..., w_{j_{h-1}i}\}$ that is $\leq c_1(s/h)^{-c_0d_0}$. Define W_{1n} as follows: $w_{ij,1} = w_{ij}$ when $w_{ij} \leq c_1(s/h)^{-c_0d_0}$; $w_{ij,1} = 0$ when $w_{ij} > c_1(s/h)^{-c_0d_0}$. W_{2n} is defined by $w_{ij,2} = w_{ij} - w_{ij,1}$. Thus every element in W_{2n} is either 0 or $> c_1(s/h)^{-c_0d_0}$. Hence,

$$\begin{aligned}\sum_{\mathscr{P}(2)} w_{ij_1}w_{j_1j_2}...w_{j_{h-1}i} &\leq [(W_{1n} + W_{2n})^h]_{ii} - (W_{2n}^h)_{ii} \leq c_1(s/h)^{-c_0d_0}\sum_{k=0}^{h-1}\|W_{2n}\|_\infty^k\|W_n\|_1^{h-k-1} \\ &\leq \left[c_1h^{c_0d_0}\sum_{k=0}^{h-1}\|W_n\|_\infty^k\tilde{c}^*(h-k-1)c_w^{h-k-1}\right]s^{-c_0d_0} \\ &\leq \left[\tilde{c}^*c_1c_w^{h-1}h^{c_0d_0}\sum_{k=0}^{h-1}(h-k-1)\right]s^{-c_0d_0} \leq \left[\tilde{c}^*c_1c_w^{h-1}h^{c_0d_0+1}(h-1)/2\right]s^{-c_0d_0} \\ &= \tilde{c}_{2h}^*s^{-c_0d_0} \text{ for some constant } \tilde{c}_{2h}^* > 0,\end{aligned}$$

where the second and third inequalities follow from Lemma A.3 and Lemma 1 in Xu and Lee (2015) respectively. Denote $g_{i,h} \equiv \mathcal{L}_i^{(h)}\left(\zeta^0, z_i, l_1, \ldots, l_{p_0}\right)(W_n^h)_{ii}$, we have

$$
\begin{aligned}
\|g_{i,h} - \mathrm{E}[g_{i,h}|\mathcal{F}_i(s)]\|_2 \leq & \sum_{\mathcal{P}(1)} w_{ij_1} w_{j_1 j_2} \ldots w_{j_{h-1} i} \cdot \|\mathcal{L}_i^{(h)}\left(\zeta^0, z_i, l_1, \ldots, l_{p_0}\right) - \mathrm{E}[\mathcal{L}_i^{(h)}\left(\zeta^0, z_i, l_1, \ldots, l_{p_0}\right)|\mathcal{F}_i(s)]\|_2 \\
& + \sum_{\mathcal{P}(2)} w_{ij_1} w_{j_1 j_2} \ldots w_{j_{h-1} i} \\
& \leq 2 c_w^h b_{\mathcal{L}_i^h} + \tilde{c}_{2h}^* s^{-c_0 d_0} \leq \tilde{c}_i^{*(h)} s^{-c_0 d_0} \text{ for some constant } \tilde{c}_i^{*(h)} > 0,
\end{aligned}
$$

where the second inequality holds because the partial derivatives (and their combinations) in $\mathcal{L}_i^{(h)}\left(\zeta^0, z_i, l_1, \ldots, l_{p_0}\right)$ are bounded under Assumption 4(iii) in the main draft, then $\mathcal{L}_i^{(h)}\left(\zeta^0, z_i, l_1, \ldots, l_{p_0}\right)$ is bounded (suppose by $b_{\mathcal{L}_i^h}$). The NED property of $\ln|I_n - \Lambda(\zeta, Z_n) W_n|$ follows from

$$
\begin{aligned}
& \|\ln|I_n - \Lambda(\zeta, Z_n) W_n| - \mathrm{E}[\ln|I_n - \Lambda(\zeta, Z_n) W_n| |\mathcal{F}_i(s)]\|_2 \\
& \leq \sum_{l=1}^{\infty} \sum_{l_1+\cdots+l_{p_0}=l} \frac{\zeta_1^{l_1} \cdots \zeta_{p_0}^{l_{p_0}}}{l_1! \cdots l_{p_0}!} \sum_{i=1}^{n} [\|g_{i,2} - \mathrm{E}[g_{i,2}|\mathcal{F}_i(s)]\|_2 + \cdots + \|g_{i,l} - \mathrm{E}[g_{i,l}|\mathcal{F}_i(s)]\|_2] \\
& \leq \sum_{l=1}^{\infty} \sum_{l_1+\cdots+l_{p_0}=l} \frac{\zeta_1^{l_1} \cdots \zeta_{p_0}^{l_{p_0}}}{l_1! \cdots l_{p_0}!} \sum_{i=1}^{n} \left(\tilde{c}_i^{*(2)} s^{-c_0 d_0} + \cdots + \tilde{c}_i^{*(l)} s^{-c_0 d_0}\right) \\
& \leq s^{-c_0 d_0} \sum_{l=1}^{\infty} \sum_{l_1+\cdots+l_{p_0}=l} \frac{\zeta_1^{l_1} \cdots \zeta_{p_0}^{l_{p_0}}}{l_1! \cdots l_{p_0}!} \sum_{i=1}^{n} (\tilde{c}_i^{*(2)} + \cdots + \tilde{c}_i^{*(l)}) \\
& \leq \tilde{C}_0^* s^{-c_0 d_0} \text{ for some constant } \tilde{C}_0^* > 0.
\end{aligned}
$$

As

$$
\begin{aligned}
& |\ln|I_n - \Lambda(\zeta, Z_n) W_n|| \\
& = \left| \sum_{l=1}^{\infty} \sum_{l_1+\cdots+l_{p_0}=l} \frac{\zeta_1^{l_1} \cdots \zeta_{p_0}^{l_{p_0}}}{l_1! \cdots l_{p_0}!} \sum_{i=1}^{n} \left[\mathcal{L}_i^{(2)}(\zeta^0, z_i, l_1, \ldots, l_{p_0})(W_n^2)_{ii} + \cdots + \mathcal{L}_i^{(l)}(\zeta^0, z_i, l_1, \ldots, l_{p_0})(W_n^l)_{ii}\right] \right| \\
& \leq \sum_{l=1}^{\infty} \sum_{l_1+\cdots+l_{p_0}=l} \frac{\zeta_1^{l_1} \cdots \zeta_{p_0}^{l_{p_0}}}{l_1! \cdots l_{p_0}!} \sum_{i=1}^{n} \left[\left|\mathcal{L}_i^{(2)}(\zeta^0, z_i, l_1, \ldots, l_{p_0})\right| \|W_n\|_\infty^2 + \cdots + \left|\mathcal{L}_i^{(l)}(\zeta^0, z_i, l_1, \ldots, l_{p_0})\right| \|W_n\|_\infty^l\right] \\
& \leq \sum_{l=1}^{\infty} \sum_{l_1+\cdots+l_{p_0}=l} \frac{\zeta_1^{l_1} \cdots \zeta_{p_0}^{l_{p_0}}}{l_1! \cdots l_{p_0}!} \sum_{i=1}^{n} (b_{\mathcal{L}_i^2} c_w^2 + \cdots + b_{\mathcal{L}_i^l} c_w^l) < \infty,
\end{aligned}
$$

By Minkowski's inequality, we have the uniform L_p boundedness $\|\ln|I_n - \Lambda(\zeta, Z_n) W_n|\|_p < \infty$.

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

$$
\begin{aligned}
& \frac{1}{n}(\ln|I_n - \Lambda(\zeta_1, Z_n) W_n| - \ln|I_n - \Lambda(\zeta_2, Z_n) W_n|) \\
& = \left|(\zeta_{1,1} - \zeta_{1,2}) \frac{1}{n} \mathrm{tr}(G_n(\bar{\zeta}_1)) + \cdots + (\zeta_{p_0,1} - \zeta_{p_0,2}) \frac{1}{n} \mathrm{tr}(G_n(\bar{\zeta}_{p_0}))\right| \\
& \leq |\zeta_{1,1} - \zeta_{1,2}| \cdot \tilde{C}_1^* + \cdots + |\zeta_{p_0,1} - \zeta_{p_0,2}| \cdot \tilde{C}_{p_0}^*
\end{aligned}
$$

where ζ is between ζ_1 and ζ_2, $G_n(\zeta_\iota) = \text{diag}(\frac{\partial\lambda(\zeta,z_i)}{\partial\zeta_\iota})W_n$ for $\iota = 1,...,p_0$, and $\tilde{C}^*_\iota$ is a constant. The inequality holds because $\sup_\zeta \|G_n(\zeta_\iota)\|_\infty < \infty$. This completes the argument for the desired uniform convergence result. □

The two claims below are the NED properties for relevant statistics under the endogenous heterogeneity specification.

Claim S.1. *Let $t_i(m)$ be the ith element of the vector $\mathcal{D}_n[\Lambda(\zeta, Z_n)W_n]^m \boldsymbol{\varphi}^*_n a$, where $\mathcal{D}_n$ is a diagonal matrix with bounded elements, which can be an identity matrix I_n or a diagonal matrix of some globally bounded functions of ζ and z, i.e., $\text{diag}\left(\boldsymbol{d}(\zeta, z_i)\right)$, with $b_{\mathcal{D}} = \sup_{\zeta,z}|\boldsymbol{d}(\zeta, z)|$; $\varphi^*_i = f_i(v_i, X_n, Z_n)$ and a is any conformable vector of constants. Under Assumptions 3, 4(i), 4(iii), and (5) in the main draft, suppose $\sup_{n.i}\|\varphi^*_i\|_p < \infty$, then $\sup_{n.i}\|t_{i,n}(m)\|_p < \infty$ and $\sup_{n.i}\|t_i(m) - \text{E}(t_i(m)|\mathcal{F}_i(s))\|_p \leq C_{a\varphi} m^{c_0 d_0+2}(b_\lambda c_w)^m s^{(1-c_0)d_0}$ with $C_{a\varphi}$ being a finite constant.*

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

$$|t_i(m)| = \left|\sum_{j=1}^n \boldsymbol{d}_i(\zeta, z_i)\left([\Lambda(\zeta, Z_n)W_n]^m\right)_{ij} \varphi^*_j a\right| \leq \sum_{j=1}^n |b_{\mathcal{D}}| \left|\left([\Lambda(\zeta, Z_n)W_n]^m\right)_{ij}\right| |\varphi^*_j a| \leq b_{\mathcal{D}} \sum_{j=1}^n \left(|b_\lambda W_n|^m\right)_{ij} |\varphi^*_j a|,$$

the second inequality holds by Assumption 4(iii) in the main draft. By Minkowski's inequality, $\|t_i(m)\|_p \leq b_{\mathcal{D}} \sum_{j=1}^n \left(|b_\lambda W_n|^m\right)_{ij} \|\varphi^*_j a\|_p < \infty$. By Proposition 1 and its proof in Jenish and Prucha (2012),

$$\sup_{n.i}\|t_i(m) - \text{E}(t_i(m)|\mathcal{F}_i(s))\|_p \leq b_{\mathcal{D}} \cdot \sup_{n.i} \sum_{j:d_{ij}>s} \left(|b_\lambda W_n|^m\right)_{ij} \sup_{n.j}\|\varphi^*_j a\|_p.$$

Under Assumption 5 in the main draft, by Eq.(C.1) in the proof of Claim C.1.6 in Qu and Lee (2015), $(W^m_n)_{ij} \leq c_2 m^{c_0 d_0+2} c^m_w d^{-c_0 d_0}_{ij}$ for some constant $c_2 > 0$. Then, for any $j \neq i$, $(|b_\lambda W_n|^m)_{ij} \leq c_2 m^{c_0 d_0+2} b^m_\lambda c^m_w d^{-c_0 d_0}_{ij}$. As $|\{j : \mathcal{K} \leq d_{ij} < \mathcal{K}+1\}| \leq c_3 \mathcal{K}^{d_0-1}$ for some constant $c_3 > 0$ by Lemma A.1 in Jenish and Prucha (2009), when s is large enough,

$$\begin{aligned}
&\sup_{n.i} \sum_{j:d_{ij}>s} \left(|b_\lambda W_n|^m\right)_{ij} \leq c_2 m^{c_0 d_0+2}(b_\lambda c_w)^m \sup_{n.i} \sum_{\mathcal{K}=[s]}^\infty \sum_{j:\mathcal{K}\leq d_{ij}<\mathcal{K}+1} d^{-c_0 d_0}_{ij} \\
&\leq c_2 c_3 m^{c_0 d_0+2}(b_\lambda c_w)^m \sum_{\mathcal{K}=[s]}^\infty \mathcal{K}^{d_0-1}\mathcal{K}^{-c_0 d_0} \leq c_2 c_3 m^{c_0 d_0+2}(b_\lambda c_w)^m \sum_{\mathcal{K}=[s]}^\infty (\mathcal{K}+1)^{d_0-1}[(\mathcal{K}+1)/2]^{-c_0 d_0} \\
&\leq c_2 c_3 m^{c_0 d_0+2}(b_\lambda c_w)^m 2^{c_0 d_0} \int_s^\infty x^{-c_0 d_0+d_0-1} dx = c_2 c_3 m^{c_0 d_0+2}(b_\lambda c_w)^m 2^{c_0 d_0}(c_0 d_0 - d_0)^{-1} s^{(1-c_0)d_0},
\end{aligned}$$

which implies that $\sup_{n.i}\|t_i(m) - \text{E}(t_i(m)|\mathcal{F}_i(s))\|_p \leq c_2 c_3 c_{a\varphi} b_{\mathcal{D}} m^{c_0 d_0+2}(b_\lambda c_w)^m 2^{c_0 d_0}(c_0 d_0 - d_0)^{-1} s^{(1-c_0)d_0}$, where $c_{ap} = \sup_{n.j}\|\varphi^*_j a\|_p$. □

Claim S.2. *Let $\tilde{g}_i(m) = e'_i \mathcal{D}_{1n} \tilde{G}^m_n(\zeta) \boldsymbol{\varphi}^*_n a$, where $e_i = (0, ..., 0, 1, 0, ..., 0)'$, $\tilde{G}_n(\zeta) = \mathcal{D}_{2n} W_n [I_n - \Lambda(\zeta, Z_n)W_n]^{-1}$ with $\mathcal{D}_{1n} = \text{diag}(\boldsymbol{d}_1(\zeta, z_i))$ and $\mathcal{D}_{2n} = \text{diag}(\boldsymbol{d}_2(\zeta, z_i))$ being two diagonal matrices of some globally bounded functions of ζ and z with $b_{\mathcal{D}_1} = \sup_{\zeta,z}|\boldsymbol{d}_1(\zeta, z)|$ and $b_{\mathcal{D}_2} = \sup_{\zeta,z}|\boldsymbol{d}_2(\zeta, z)|$. Under Assumption 3, 4(i), 4(iii), and 5 in the main draft, suppose $\sup_{n.i}\|\varphi^*_i\|_p < \infty$, then $\sup_{n.i}\|\tilde{g}_i(m)\|_p < \infty$ and $\sup_{n.i}\|\tilde{g}_i(m) - \text{E}(\tilde{g}_i(m)|\mathcal{F}_i(s))\|_p \leq C_{a\varphi m} s^{(1-c_0)d_0}$ with $C_{a\varphi m}$ being a finite constant.*

Proof. By the proof of Claim C.1.7 in Qu and Lee (2015), $\tilde{G}_n^m(\zeta) = [I_n - \Lambda(\zeta, Z_n)W_n]^{-m}(\mathcal{D}_{2n}W_n)^m = \sum_{l=0}^{\infty} C_l^{l+m-1}[\Lambda(\zeta, Z_n)W_n]^l(\mathcal{D}_{2n}W_n)^m \leq b_{\mathcal{D}_2}^m \sum_{l=0}^{\infty} C_l^{l+m-1} b_\lambda^l W_n^{l+m}$, where C_l^{l+m-1} is a binomial coefficient. Using the result for $t_i(l+m)$ in Claim S.1, we have

$$\sup_{n.i} \|\tilde{g}_i(m)\|_p \leq b_{\mathcal{D}_1} b_{\mathcal{D}_2}^m b_\lambda^{-m} \sum_{l=0}^{\infty} C_l^{l+m-1} \sup_{n.i} \|t_i(l+m)\|_p < \infty,$$

and

$$\begin{aligned}
\sup_{n.i}\|\tilde{g}_i(m) - \mathrm{E}(\tilde{g}_i(m)|\mathcal{F}_i(s))\|_p &\leq b_{\mathcal{D}_1} b_{\mathcal{D}_2}^m b_\lambda^{-m} \sum_{l=0}^{\infty} C_l^{l+m-1} \sup_{n.i} \|t_i(l+m) - \mathrm{E}\left(t_i(l+m)|\mathcal{F}_i(s)\right)\|_p \\
&\leq b_{\mathcal{D}_1} b_{\mathcal{D}_2}^m C_{a\varphi} \sum_{l=0}^{\infty} b_\lambda^l c_w^{l+m} (l+m)^{m+1+c_0 d_0} s^{(1-c_0)d_0} \leq C_{a\varphi m} s^{(1-c_0)d_0}.
\end{aligned}$$

□

Let $\tilde{\mathcal{M}}_n = \tilde{A}_n' \tilde{B}_n$, where $\tilde{A}_n$ and $\tilde{B}_n$ are either $\mathcal{D}_n \left[\Lambda\left(\zeta, Z_n\right) W_n\right]^{m_1}$ or $\tilde{G}_n^{m_2}(\zeta)$ with m_1 and m_2 being finite non-negative integers. The NED property of the statistic $a'\boldsymbol{\varphi}_n^{*\prime}\tilde{\mathcal{M}}_n\boldsymbol{\varphi}_n^* b$ for some constant vectors a and b with φ_i^* as the basis for the NED is established in Claim S.1-S.2. Then we have the following LLN, ULLN and CLT based on the asymptotic inference under NED.

Proposition S.1. *Under Assumptions 1-5 and 8 in the main draft (i) suppose* $\sup_{n.i}\|\varphi_i^*\|_4 < \infty$*, then* $\frac{1}{n}\mathrm{E}|a'\boldsymbol{\varphi}_n^{*\prime}\tilde{\mathcal{M}}_n\boldsymbol{\varphi}_n^* b| = O(1)$ *and* $\frac{1}{n}[a'\boldsymbol{\varphi}_n^{*\prime}\tilde{\mathcal{M}}_n\boldsymbol{\varphi}_n^* b - \mathrm{E}(a'\boldsymbol{\varphi}_n^{*\prime}\tilde{\mathcal{M}}_n\boldsymbol{\varphi}_n^* b)] = o_p(1)$*.*
(ii) suppose $\sup_{n.i}\|\varphi_i^*\|_4 < \infty$*, and denote* $\varphi_n^*(\theta) = f_i(v_i, Z_n, X_n, \theta)$ *with* θ *entering* f_i *polynomially, then* $\frac{1}{n}a'\varphi_n^*(\theta)'\tilde{G}_n^{m_1}(\zeta)'\mathcal{D}_{1n}^2\tilde{G}_n^{m_2}(\zeta)\varphi_n^*(\theta)b$ *(and* $\frac{1}{n}a'\varphi_n^*(\theta)'\tilde{G}_n^{m_1}(\zeta)'\tilde{G}_n^{m_2}(\zeta)\varphi_n^*(\theta)b$*) is stochastic equicontinuous and* $\sup_{\theta\in\Theta}\frac{1}{n}|a'\varphi_n^*(\theta)'\tilde{G}_n^{m_1}(\zeta)'\mathcal{D}_{1n}^2\tilde{G}_n^{m_2}(\zeta)\varphi_n^*(\theta)b - \mathrm{E}(a'\varphi_n^*(\theta)'\tilde{G}_n^{m_1}(\zeta)'\mathcal{D}_{1n}^2\tilde{G}_n^{m_2}(\zeta)\varphi_n^*(\theta)b)| = o_p(1)$*.*
(iii) suppose $\sup_{n.i}\|\varphi_i^*\|_{4+\mathfrak{d}} < \infty$ *for some* $\mathfrak{d} > 0$*, and* $\inf_n \frac{1}{n}\sigma_{\mathcal{S}_n}^2 > 0$*, where* $\mathcal{S}_n = \sum_{j=1}^m [a_j'\boldsymbol{\varphi}_n^{*\prime}\tilde{\mathcal{M}}_j\boldsymbol{\varphi}_n^* b_j - \mathrm{E}(a_j'\boldsymbol{\varphi}_n^{*\prime}\tilde{\mathcal{M}}_j\boldsymbol{\varphi}_n^* b_j)] = \sum_{i=1}^n \boldsymbol{s}_i$ *and* $\sigma_{\mathcal{S}_n}^2 = \mathrm{Var}(\sum_{i=1}^n \boldsymbol{s}_i)$*, then* $\mathcal{S}_n/\sigma_{\mathcal{S}_n} \xrightarrow{d} N(0,1)$*.*

Proof. We show the stochastic equicontinuity of $\frac{1}{n}a'\varphi_n^{*\prime}\tilde{G}_n^{m_1}(\zeta)'\mathcal{D}_{1n}^2\tilde{G}_n^{m_2}(\zeta)\varphi_n^* b$ as all the other results follow the similar arguments in the proofs of Proposition 1, Corollary 1 and Proposition 2 in Qu and Lee (2015) given the results of Claim S.1 and Claim S.2. By the mean value theorem,

$$\begin{aligned}
&|a'\varphi_n^{*\prime}\tilde{G}_n^{m_1}(\zeta_1)'\mathcal{D}_{1n}^2\tilde{G}_n^{m_2}(\zeta_1)\varphi_n^* b - a'\varphi_n^{*\prime}\tilde{G}_n^{m_1}(\zeta_2)'\mathcal{D}_{1n}^2\tilde{G}_n^{m_2}(\zeta_2)\varphi_n^* b| \\
&\quad = |(\zeta_{1,1} - \zeta_{1,2})a'\varphi_n^{*\prime}A_n(\bar{\zeta}_1)\varphi_n^* b + \cdots + (\zeta_{p_0,1} - \zeta_{p_0,2})a'\varphi_n^{*\prime}A_n(\bar{\zeta}_{p_0})\varphi_n^* b| \\
&\quad \leq |\zeta_{1,1} - \zeta_{1,2}|(a'\varphi_n^{*\prime}\varphi_n^* a)^{\frac{1}{2}}(b'\varphi_n^{*\prime}A_n(\bar{\zeta}_1)'A_n(\bar{\zeta}_1)\varphi_n^* b)^{\frac{1}{2}} + \cdots + |\zeta_{p_0,1} - \zeta_{p_0,2}|(a'\varphi_n^{*\prime}\varphi_n^* a)^{\frac{1}{2}}(b'\varphi_n^{*\prime}A_n(\bar{\zeta}_{p_0})'A_n(\bar{\zeta}_{p_0})\varphi_n^* b)^{\frac{1}{2}} \\
&\quad \leq |\zeta_{1,1} - \zeta_{1,2}|(a'\varphi_n^{*\prime}\varphi_n^* a)^{\frac{1}{2}}(b'\varphi_n^{*\prime}\varphi_n^* b)^{\frac{1}{2}}[\mu_{max}(A_n(\bar{\zeta}_1)'A_n(\bar{\zeta}_1))]^{\frac{1}{2}} + \cdots \\
&\qquad + |\zeta_{p_0,1} - \zeta_{p_0,2}|(a'\varphi_n^{*\prime}\varphi_n^* a)^{\frac{1}{2}}(b'\varphi_n^{*\prime}\varphi_n^* b)^{\frac{1}{2}}[\mu_{max}(A_n(\bar{\zeta}_{p_0})'A_n(\bar{\zeta}_{p_0}))]^{\frac{1}{2}} \\
&\quad \leq |\zeta_{1,1} - \zeta_{1,2}|(a'\varphi_n^{*\prime}\varphi_n^* a)^{\frac{1}{2}}(b'\varphi_n^{*\prime}\varphi_n^* b)^{\frac{1}{2}}\left(\sup_{\zeta\in\Theta_\zeta}\|A_n'(\zeta_1)A_n(\zeta_1)\|_\infty\right)^{\frac{1}{2}} + \cdots \\
&\qquad + |\zeta_{p_0,1} - \zeta_{p_0,2}|(a'\varphi_n^{*\prime}\varphi_n^* a)^{\frac{1}{2}}(b'\varphi_n^{*\prime}\varphi_n^* b)^{\frac{1}{2}}\left(\sup_{\zeta\in\Theta_\zeta}\|A_n'(\zeta_{p_0})A_n(\zeta_{p_0})\|_\infty\right)^{\frac{1}{2}}
\end{aligned} \tag{S.30}$$

where $\bar{\zeta}$ is between ζ_1 and ζ_2, $A_n(\zeta_\iota) = \tilde{G}_n^{m_1}(\zeta)'[m_1 G_n'(\zeta_\iota)\mathcal{D}_{1n}^2 + m_2\mathcal{D}_{1n}^2 G_n(\zeta_\iota)]\tilde{G}_n^{m_2}(\zeta)$ with $G_n(\zeta_\iota) = \text{diag}(\frac{\partial\lambda(\zeta,z_i)}{\partial\zeta_\iota})W_n S_n^{-1}(\zeta)$ for $\iota = 1,...,p_0$ and $\mu_{max}(\cdot)$ is the largest eigenvalue of the matrix inside. The first inequality is from the Cauchy-Schwarz inequality, the second inequality holds as $A_n\left(\bar{\zeta}_\iota\right)' A_n\left(\bar{\zeta}_\iota\right)$ ($\iota = 1,...,p_0$) are non-negative definite, and the last inequality holds by spectrum radius theorem. The stochastic equicontinuity of $\frac{1}{n}a'\varphi_n^*(\theta)'\tilde{G}_n^{m_1}(\zeta)'\tilde{G}_n^{m_2}(\zeta)\varphi_n^*(\theta)b$ can be shown in a similar fashion. □

S.5 Additional Simulations

The IV-estimable endogenous heterogeneity setting with different sample sizes. In this case, we have a slightly different model $y_i = \varrho F(z_i)\sum_{j\neq i} w_{ij}y_j + x_i'\beta + v_i$ so that the IV method can be applied (as shown in Section 3.3.2), where the bounded function $F(z_i)$ is a Logistic function with the location parameter $\mu = 0$ and the scale parameter $s = 1$. To generate the data, we follow equation (39) in the main draft for the joint distribution of (x_i^*, z_i^*, v_i^*). As for regressors, we have $x_i = (x_{1,i}, x_{2,i}, x_{3,i})'$, where $x_{1,i} = 1$, $x_{2,i} \sim N(0,1)$, and $x_{3,i} = R_3^{-1}(\Phi(x_{3,i}^*))$, with R_3 being the CDF for the Exponential distribution with $\mu = 2$ [2]. $z_i = H^{-1}(\Phi(z_i^*)) = \Phi^{-1}(\Phi(z_i^*)) = z_i^*$. The spatial weight matrix is W_n^d in the main text. Table S.1 displays that our estimator gives precise estimates in this setting with different sample sizes.

MLE with group fixed effect. In this experiment, we provide estimates for the endogenous heterogeneity model with group fixed effects as in Section 5 in the main draft. We stick to the setting for endogenous heterogeneity in Table 2 (main draft) and divide the $n = 144$ cross-sectional units into 15 fixed groups. Group fixed effect c_g is uncorrelated with any observable and is distributed exponentially [3]. Table S.2 evaluates two estimators – Raw refers to θ_{ML} and Correction is $\hat{\theta}_{ML}^c$ in equation (38) in the main draft. In this specific example, the bias in θ_{ML} is already small, which does not leave enough room for improvement. Even so, the correction in θ_{ML}^c pulls the estimates to the true values in three or four decimal places and provides an improvement in the coverage of σ_v. Therefore, we still recommend the use of correction for robustness.

2. This is no longer normally distributed since $F(z_i)$ is highly correlated with z_i, which may create identification problem as in the cases with linear regressors.

3. Results in Table S.2 are robust to the distribution of fixed effects. We omitted further simulations due to limited space.

			n = 49		n = 144		n = 361		n = 529		n = 1024	
	True		MLE	IV	MLE	IV	MLE	IV	MLE	IV	MLE	IV
β_0	1	Mean	0.9454	0.9635	0.9868	0.9935	0.9937	0.9963	0.9953	0.9973	0.9953	0.9961
		Std	(0.2614)	(0.2656)	(0.1417)	(0.1438)	(0.0942)	(0.0948)	(0.0726)	(0.0729)	(0.0532)	(0.0536)
			[0.2354]	[0.2363]	[0.1332]	[0.1345]	[0.0797]	[0.0847]	[0.0632]	[0.0701]	[0.0461]	[0.0503]
		Coverage	0.9490	0.9520	0.9510	0.9540	0.9500	0.9570	0.9500	0.9470	0.9480	0.9440
β_1	4	Mean	3.9985	3.9983	3.9995	3.9994	4.0018	4.0017	3.9998	3.9997	3.9996	3.9995
		Std	(0.1234)	(0.1233)	(0.0713)	(0.0714)	(0.0445)	(0.0445)	(0.0354)	(0.0354)	(0.0250)	(0.0250)
			[0.0974]	[0.1196]	[0.0681]	[0.0688]	[0.0438]	[0.0432]	[0.0369]	[0.0356]	[0.0255]	[0.0255]
		Coverage	0.9440	0.9440	0.9530	0.9530	0.9520	0.9520	0.9460	0.9460	0.9490	0.9500
β_2	-2	Mean	-2.0011	-1.9987	-2.0032	-2.0023	-2.0005	-2.0001	-2.0003	-2.0000	-1.9996	-1.9995
		Std	(0.0817)	(0.0823)	(0.0435)	(0.0436)	(0.0270)	(0.0271)	(0.0226)	(0.0226)	(0.0158)	(0.0158)
			[0.0451]	[0.0762]	[0.0318]	[0.0428]	[0.0214]	[0.0268]	[0.0189]	[0.0220]	[0.0135]	[0.0158]
		Coverage	0.9530	0.9550	0.9420	0.9380	0.9420	0.9460	0.9510	0.9470	0.9510	0.9530
ϱ	0.5	Mean	0.4828	0.4948	0.4940	0.4984	0.4981	0.4998	0.4989	0.5002	0.4995	0.5001
		Std	(0.0991)	(0.1041)	(0.0525)	(0.0546)	(0.0338)	(0.0343)	(0.0270)	(0.0278)	(0.0191)	(0.0194)
			[0.0724]	[0.0936]	[0.0361]	[0.0524]	[0.0281]	[0.0327]	[0.0232]	[0.0272]	[0.0177]	[0.0196]
		Coverage	0.9460	0.9470	0.9520	0.9520	0.9440	0.9510	0.9500	0.9470	0.9510	0.9530
ρ_{vz}	0.5	Mean	0.4709	-	0.4899	-	0.4955	-	0.4959	-	0.4983	-
		Std	(0.1058)	-	(0.0587)	-	(0.0380)	-	(0.0317)	-	(0.0229)	-
			[0.1103]	-	[0.0654]	-	[0.0415]	-	[0.0343]	-	[0.0247]	-
		Coverage	0.9420	-	0.9430	-	0.9480	-	0.9510	-	0.9450	-
σ_v	1	Mean	0.9579	-	0.9862	-	0.9944	-	0.9956	-	0.9977	-
		Std	(0.1069)	-	(0.0617)	-	(0.0392)	-	(0.0318)	-	(0.0230)	-
			[0.0968]	-	[0.0581]	-	[0.0370]	-	[0.0306]	-	[0.0220]	-
		Coverage	0.9370	-	0.9410	-	0.9510	-	0.9460	-	0.9540	-

Table S.1: Estimates from the IV-estimable setting under different sample sizes.

		True	Mean	Std	Coverage	p10	p30	p50	p70	p90
β_1	Raw	4	3.9965	(0.0738)	0.9410	-0.0957	-0.0398	-0.0029	0.0309	0.0901
	Correction		3.9959	(0.0738) [0.0705]	0.9410	-0.0958	-0.0400	-0.0032	0.0305	0.0897
β_2	Raw	-2	-2.0067	(0.0554)	0.9570	-0.0787	-0.0376	-0.0074	0.0203	0.0667
	Correction		-2.0064	(0.0554) [0.0342]	0.9570	-0.0788	-0.0375	-0.0071	0.0212	0.0667
ρ	Raw	0.8	0.7877	(0.0589)	0.9350	-0.0908	-0.0398	-0.0099	0.0172	0.0596
	Correction		0.7943	(0.0589) [0.0540]	0.9350	-0.0847	-0.0333	-0.0036	0.0234	0.0665
ϱ	Raw	0.5	0.5396	(0.1583)	0.9420	-0.1538	-0.0458	0.0275	0.1120	0.2480
	Correction		0.5341	(0.1561) [0.1525]	0.9420	-0.1563	-0.0499	0.0228	0.1058	0.2369
ρ_{vz}	Raw	0.5	0.4937	(0.0941)	0.9500	-0.1281	-0.0447	0.0041	0.0459	0.1021
	Correction		0.4781	(0.0933) [0.0656]	0.9430	-0.1427	-0.0606	-0.0124	0.0292	0.0863
σ_v	Raw	1	0.9481	(0.0944)	0.9190	-0.1673	-0.1051	-0.0569	-0.0051	0.0639
	Correction		0.9795	(0.0932) [0.0578]	0.9470	-0.1349	-0.0716	-0.0253	0.0263	0.0941

Table S.2: Estimates from the endogenous heterogeneity setting with group fixed effects. n = 144.

References

Haschka, R. 2022. "Handling endogenous regressors unsing copulas: a generalization to linear panel models with fixed effects and correlated regressors." *Journal of Marketing Research* 59 (4): 860–881.

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.

———. 2012. "On spatial processes and asymptotic inference under near-epoch dependence." *Journal of Econometrics* 170 (1): 178–190.

Kelejian, H., and I. Prucha. 1998. "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance." *Journal of Real Estate Finance and Economics* 17:99–121.

Liu, H., F. Han, M. Yuan, J. Lafferty, and L. Wasserman. 2012. "High-dimensional semiparametric Gaussian copula graphical models." *The Annals of Statistics* 30 (4): 2293–2326.

Park, S., and S. Gupta. 2012. "Handling Endogenous Regressors by Joint Estimation Using Copulas." *Marketing Science* 31 (4): 567–586.

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.

Rothenberg, T. 1971. "Identification in parametric models." *Econometrica* 39 (3): 577–591.

Sklar, A. 1959. "Fonctions de. répartition n dimensions et leurs marges." *Publications de l′Institut de Statistique de L′Universit de Paris* 8:229–231.

Xu, X., and L. Lee. 2015. "Maximum likelihood estimation of a spatial autoregressive Tobit model." *Journal of Econometrics* 188 (1): 264–280.

Yang, F., Y. Qian, and H. Xie. 2022. "Addressing Endogeneity Using a Two-stage Copula Generated Regressor Approach." NBER Working Paper 29708, https://www.nber.org/papers/w29708, July.