Addressing endogeneity issues in a SAR model using copulas

Yanli Lin[*], Yichun Song[†]

October 15, 2022

Abstract

We provide a new copula method to tackle possible endogeneity issues in a spatial autoregressive (SAR) model, which might originate from an endogenous spatial weight matrix, endogenous heterogeneity specification, or endogenous regressors. Model specifications and estimations for the three variants of a SAR model are presented. Using copula endogeneity correction technique, we propose 3-stage estimation methods and establish their consistency and asymptotic normality. Monte Carlo experiments are performed to investigate finite sample performance of the estimators. We then apply our method to an empirical study of spatial spillovers in regional productivity with endogenous spatial weights constructed by the proximity of a "meaningful" socioeconomic characteristic - years of education.

Keywords: Spatial autoregressive model, Endogenous spatial weight matrix, Endogenous heterogeneity, Endogenous regressors, Copula method
JEL classification: C31, C51

1 Introduction

The ways in which the outcomes of interconnected spatial units influence each other are usually referred to as spillover effects, which are usually captured by the widely studied spatial autoregressive (SAR) model as it has a game structure and can be regarded as a reaction function. Early research of the traditional SAR model can be found in Cliff and Ord (1973), Ord (1975), Anselin (1988), Kelejian and Prucha (2001), and Lee (2004, 2007), where various estimation methods, such as maximum likelihood estimation (MLE), instrumental variable (IV) methods, and generalized method of moments (GMM), are provided. Some basic assumptions are imposed on the traditional SAR model, for instance, linear model structure, an exogenous spatial weight matrix, a scalar spatial coefficient, exogenous regressors, etc., which have been relaxed in recent literature in order to meet the needs of empirical studies. Typical examples are a SAR model with a nonlinear transformation of the dependent variable for "share data"(Xu and Lee, 2015a), a SAR Tobit model for censored or binary data (Xu and Lee, 2015b), a SAR model with an endogenous spatial weight matrix constructed by "economic distance" (Qu and Lee, 2015), its further extensions to a panel data setting (Qu et al., 2017) and the construction

[*]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.

of endogenous spatial weights by bilateral variables (Qu et al., 2021), nonparametric smooth coefficient SAR models with heterogeneous "reaction parameters" (Malikow and Sun, 2017), higher-order SAR model with an endogenous regression component (Gupta and Robinson, 2015), and so on.

We consider specifications and estimations of three variants of a SAR model which might have endogeneity issues caused by observed endogenous variables. The first variant has an endogenous spatial weight matrix [1], the second variant involves endogenous heterogeneity with a parametric nonlinear spatial interaction term, and the third one contains endogenous regressors, where endogenous variables enter the SAR model in a nonlinear manner in the first and second variants, but they are linear components in the last variant. One common solution to handle the endogneity issues in a SAR model is the control function approach proposed (Qu and Lee, 2015) by specifying two sets of equations: one for the SAR outcome and the other for the endogenous variables, and assuming the source of endogeneity comes from the correlation between the error terms in the stated equations. As detailed in Section 2, unique structure and correct model specification (for instance, without omitted variables) for the endogenous variables equation are required to guarantee model identification and unbiased estimates. Therefore, we directly characterize the entire dependence structure among the error term, the endogenous variables, and the exogenous variables (independent of the error term but can be correlated with the endogenous variables) using a parametric copula so that we don't need to specify the additional set of equation for the endogenous variables. The copula endogeneity correction technique to handle endogenous regressors in a linear regression model and serve as an instrument-free method can be found in the seminal work of Park and Gupta (2012), Haschka (2022) and Yang et al. (2022). We extend the basic idea of Yang et al. (2022) to spatial settings and propose three-stage estimation methods for the three variants of a SAR model because Park and Gupta (2012)'s original approach leads to asymptotic bias[2] and that of Haschka (2022) has a restrictive assumption[3].

This paper makes four advances compared to the existing literature. First, the second variant of a SAR model, together with a few other papers, for instance Xu and Lee (2015a, 2015b) and Rabovič and Čížek (2022), add to the literature on specific parametric nonlinear spatial models. Similar to the motivation and application of the nonparametric smooth coefficient spatial autoregressive models in Malikow and Sun (2017), the parametric nonlinear interaction function can be used to capture spillover effects that vary with the own characteristics of a spatial unit, refer to their paper for potential applications. Second, we provide alternative statistical copula approach to address endogeneity issues in three variants of a SAR model and discuss the differences in their required assumptions and estimation methods, while the control function approach only focuses on the endogenity of the spatial weight matrix. When the true model setting for the set of endogenous variables equation contains some extra exogenous variables (apart from the exogenous variables included in the SAR outcome equation) for the control function approach, it's not easy to find reasonable variables that satisfy the exogeneity assumption in empirical applications, thus the extra variables are usually omitted in estimation by only adding the residuals obtained from regressing the endogenous variables on the exogenous regressors in

1. As pointed out by Pinkse and Slade (2010), endogeneity of spatial weights is challenging and waiting for good solutions, and is one of the future research directions of spatial econometrics.

2. The proofs are documented in Section S.1 in the supplement file.

3. The assumption is nonnormality of endogenous regressors.

the outcome equation as control variables to control the endogeneity issue. Although identification is not an issue when the endogenous variables enter the SAR model nonlinearly (the first[4] and second variant), it will affect that of the case when we have a linear endogenous component (third variant). Besides, the control function approach is less robust against the misspecification than our proposed copula method when the omitted variables are non-normally distributed as shown in our Monte Carlo study.

Third, the copula endogeneity correction technique has been extended from the previous applications in linear regression models to accommodate more complex spatial econometric models. Yang et al. (2022) only consider a 3-stage instrument-free ordinary least squares estimator (OLS) with controlled variables for the linear regression models, while we develop 3-stage MLE and 3-stage IV methods for our SAR models. We detect that the MLE estimators can be applied when the endogenous variables enter the SAR model in a nonlinear way (first and second variant), the IV estimators can be applied to the first variant and a special case of the second variant when the nonlinear function capturing the heterogeneous spillover effects doesn't include any parameters inside. However, there might be asymptotic bias of the MLE estimator for the third variant unless the residual variance equals 1, the IV estimator can be employed for this case but it's required as noted in Yang et al. (2022) that one of the correlated exogenous regressors should be non-normally distributed for identification. Last but not least, we provide asymptotic properties for the 3-stage estimators. As mentioned in Becker et al. (2021) and Haschka (2022), one major research area for copula correction is to provide rigorous proofs of required model identification conditions and estimation properties, which is lacking in previous studies except Yang et al. (2022) for linear regression models. We show that the sampling errors caused by the employment of estimated marginal transformations (by using the proposed estimator in Liu et al. (2012)) instead of the true marginal transformations are asymptotically negligible. Then under the regularity conditions, the consistency and asymptotic normality for the 3-stage estimators can be derived using the asymptotic inference under near-epoch dependence (NED) from Jenish and Prucha (2012).

Small sample properties of the 3-stage estimation methods are investigated by a Monte Carlo experiment design. Our proposed estimators using Gaussian copulas are shown to perform well with different sample sizes, various endogenous variables, and under other non-normal joint distributions of the marginal transformations of endogenous variables and the error term. We verify that the direct extension of Park and Gupta (2012)'s approach to a SAR model causes asymptotic bias, and that at least one of the exogenous variables should be non-normally distributed when endogenous regressors are included (Yang et al., 2022). We also compare the performance of our copula endogeneity correction method and the control function approach (Qu and Lee, 2015) under the case that the endogenous variables are generated by a nonlinear function of the corresponding transformed variables, the transformed error term, and additional exogenous variables that can serve as excluded instruments. When the excluded instruments or the exact functional form are unobserved, simulation results show that our copula estimators are still robust while the control function estimator produces biased estimates with increased standard errors and low levels of 95% coverage probability when the omitted excluded

4. This point has been mentioned in Qu and Lee (2015).

instruments are non-normally distributed.

We provide an empirical application: the regional productivity spillovers in $1,433$ subnational regions from 110 different countries. The spatial weights are determined by years of education, which is a possible endogenous economic factor. Estimate for the spatial coefficient from our copula method is closed to that from the control function approach (Qu and Lee, 2015) when the years of education variable is only part of the spatial weights. However, under the case that years of education is also added as a regressor and there exists difficulty in getting excluded instruments, which is common in empirical studies, our proposed copula method would be good alternative for the control function approach since the latter method might not be appropriate.

The rest of the paper proceeds as follows. Section 2 sets up the three variants of a SAR model with possible endogeneity issues and explains the copula method in detail. Section 3 introduces the 3-stage estimation approaches. Their consistency and asymptotic distribution are discussed in the subsequent section. Section 5 extends the SAR model to the case with group fixed effects and demonstrates a biased corrected estimator. Finite sample performance are examined by Monte Carlo (MC) simulation studies in Section 6. Section 7 applies our proposed methods to empirically study spatial spillovers of regional economic performance. Section 8 concludes the paper. Related statistics of the log pseudo-likelihood function and detailed mathematical proofs are collected in the Appendix. Additional proofs and MC results can be found in the supplementary file.

2 Model

We consider a cross-sectional SAR model for the outcome y at the location i as[5]

$$y_i = \lambda \sum_{j \neq i} w_{ij} y_j + x_i' \beta + v_i \tag{1}$$

for $i = 1, .., n$, w_{ij} measures the relative strength of linkage between location i and location j ($i \neq j$, $w_{ii} = 0$ for all i), $x_i = (x_{1,i}, ..., x_{k,i})' \in \mathbb{R}^k$ is a vector of observed exogenous variables, v_i is the error term. λ is a spatial coefficient and $\beta = (\beta_1, ..., \beta_k)'$ is a k-dimensional vector of parameters. The SAR model has the matrix form

$$Y_n = \lambda W_n Y_n + X_n \beta + V_n \tag{2}$$

where $Y_n = (y_1, ..., y_n)'$ and $V_n = (v_1, ..., v_n)'$ are n-dimensional vectors, $X_n = (x_1, ..., x_n)'$ is an $n \times k$ matrix, $W_n = (w_{ij})$ is an $n \times n$ nonnegative matrix with zero diagonals. When $S_n^{-1}(\lambda)$ exists, where $S_n(\lambda) = I_n - \lambda W_n$, the model has a reduced form,

$$Y_n = S_n^{-1}(\lambda)(X_n \beta + V_n). \tag{3}$$

In general, other than the endogenous $W_n Y_n$, there are three variants of the SAR model which might have endogeneity issues caused by observed endogenous variables Z_n, where $Z_n = (z_1, ..., z_n)'$ is

5. In the scalar form of the model, we omit the subscript n, which represents the identity of the nth individual, for brevity.

an $n \times p$ matrix with $z_i = (z_{1,i}, ..., z_{p,i})'$ being p dimensional column vectors. The first variant is a SAR model (Eq.(1)) with an endogenous spatial weight matrix (Qu and Lee, 2015), i.e., the elements of W_n are constructed by $Z_n : w_{ij} = \psi_{ij}(Z_n, d_{ij})$ for $i, j = 1, ..., n,\ i \neq j$, where $\psi(\cdot)$ is a bounded function. For example, $w_{ij} = w_{ij}^e w_{ij}^d$, where $w_{ij}^e = 1/|z_i - z_j|$ with z_i being some socioeconomic characteristics (Case et al., 1993), and w_{ij}^d is based on geographic distance. The second variant is a SAR model with endogenous heterogeneity,

$$y_i = \lambda\left(\zeta, z_i\right) \sum_{j \neq i} w_{ij} y_j + x_i' \beta + v_i \tag{4}$$

where $\lambda(\zeta, z_i)$ is a known globally bounded parameter function of the endogenous variables z_i and a vector of parameters ζ, for instance, $\lambda(\zeta, z_i) = \rho F(\varrho_0 + \varrho_1 z_{1,i} + \cdots + \varrho_p z_{p,i})$ where $F(\cdot)$ is a continuous probability function, $\zeta = (\rho, \varrho_0, ..., \varrho_p)'$ is a $p_0(= p + 2)$-dimensional vector. [6] Spatial dependepnce in house prices and institutional diffusion across space are two good examples of applying the smooth coefficient SAR model (Malikow and Sun, 2017), while it would also be meaningful to account for the endogenous nonlinear heterogeneity, which was not considered in their paper. The third variant is a SAR model with endogenous explanatory variables z_i,

$$y_i = \lambda \sum_{j \neq i} w_{ij} y_j + x_i' \beta_1 + z_i' \beta_2 + v_i \tag{5}$$

where $\beta_1 = (\beta_{1,1}, ..., \beta_{1,k})'$ and $\beta_2 = (\beta_{2,1}, ..., \beta_{2,p})'$ are k and p-dimensional vectors of parameters respectively. This is a direct extension of the linear regression model with one endogenous and one exogenous regressors in Park and Gupta (2012), Haschka (2022) and Yang et al. (2022) to the spatial setting with multiple endogenous and exogenous regressors.[7] In this paper, we focus on the first and second variants[8], where Z_n are incorporated in a SAR model nonlinearly, and shall emphasize the differences in estimation methods among the three specifications.

To handle the endogeneity issues in a SAR model, one common method is to employ the control function approach. The endogenous variables Z_n are represented as

$$Z_n = \mu(A_n) + \boldsymbol{\varepsilon}_n \tag{6}$$

where $A_n = (a_1, ..., a_n)'$ is an $n \times k_1$ matrix with $a_i = (a_{1,i}, ..., a_{k_1,i})'$ being k_1-dimensional column vectors of observed exogenous variables, $\boldsymbol{\varepsilon}_n = (\varepsilon_1, ..., \varepsilon_n)'$ is an $n \times p$ matrix of disturbances with $\varepsilon_i = (\varepsilon_{1,i}, ..., \varepsilon_{p,i})'$. $\mu(\cdot)$ is a function of A_n and unique structures are imposed on $\mu(A_n)$ for model identification. In Qu and Lee (2015), they consider a linear setting for $\mu(A_n)$, i.e., $Z_n = A_n \Pi + \boldsymbol{\varepsilon}_n$, where Π is a $k_1 \times p$ matrix of coefficients. The error terms v_i and ε_i are assumed to have a joint distribution, which implies a conditional homoskedasticity condition, and then $Z_n - A_n \Pi$ can be added

6. The smooth coefficient SAR model with unknown smooth functions of some relevant exogenous variables is investigated in Malikow and Sun (2017) using semiparametric estimation method. Here, we consider a slightly different parametric setting because our proposed copulas approach below, combined with some standard nonparametric econometric technique, for instance, local polynomial regression, can be generalized to accommodate the unkonwn smooth coefficient with endogenous variables specification.

7. See also the higher-order SAR model with an endogenous regression component in Gupta and Robinson (2015).

8. Because the main derivations for the first variant can be applied to the third one and the third one is less interesting in spatial setting.

as control variables to control the endogeneity of W_n. They only consider the nonlinearity of Z_n in W_n case, but the idea can be generalized to the other two cases (Eq.(4)-(5)). Without observing the structure of $\mu(A_n)$, there can be problem with the linear setting, for instance, the true specification for $\mu(A_n)$ might not be linear.[9] Even if we ignore this issue and implement the control function approach with a linear structure, there are still two notable concerns - identification and model specification. When Z_n enter into the SAR model in a nonlinear manner, i.e., the first and second variants, A_n and X_n are allowed to have common variables and they could even be the same as identification is not an issue (Qu and Lee, 2015), however, it's still required that the control function should be correctly specified, otherwise biased estimates could occur. In the Monte Carlo simulation section, we consider a misspecification scenario with omitted variables in A_n. Suppose A_n in the true model setting contain both X_n and some extra variables X_{2n}, but due to some reasons, for instance significant difficulties in finding such "good instrumental variables" X_{2n} that satisfy the exogeneity assumption[10], X_{2n} are omitted in the estimation. Simulation results show that regressing Z_n on X_n only and adding the error terms as control variables would not cause identification issue, but we have biased results when the omitted X_{2n} are non-normally distributed. When Z_n enter into a SAR model in a linear way, i.e., the third variant, apart from the model specification concern, A_n and X_n can not be the same, some extra exogenous variables, which might not be easy to find, are needed to satisfy the full rank assumption for identification.

We handle endogeneity issues in the aforementioned variants of a SAR model and address the above concerns of the control function approach by directly modelling the correlations among the endogenous variables z_i and the error term v_i using a parametric copula in light of Park and Gupta (2012).[11] However, as shown in Section S.1 of our supplement file and also mentioned in Yang et al. (2022), there is an implicit assumption of the original approach in Park and Gupta (2012), that is the exogenous regressors should be uncorrelated with the copula transformations of the endogenous variables, which are employed to control for endogeneity, and when this assumption is violated, there exists asymptotic bias in the estimation. When z_i are added as endogenous regressors, we need to impose one additional assumption for identification - non-normality of $z_{\iota,i}$ ($\iota = 1, ..., p$). To tackle the restrictions, following Haschka (2022), we characterize the entire dependence structure among x_i, z_i and v_i, while assuming that x_i is independent of v_i. Given the known marginal distributions of $x_{1,i}, ..., x_{k,i}$, $z_{1,i}, ..., z_{p,i}$ and v_i, denoted as $R_1(x_{1,i}), ..., R_k(x_{k,i})$, $H_1(z_{1,i}), ..., H_p(z_{p,i})$ and $G(v_i)$, [12] by

9. Or not all the function forms for the p endogenous variables are linear, i.e., $z_{1,n} = \mu_1(A_n) + \varepsilon_{1,n}, ..., z_{p,n} = \mu_p(A_n) + \varepsilon_{p,n}$, where $z_{\iota,n}$ and $\varepsilon_{\iota,n}$ are n dimensional vectors for $\iota = 1, ..., p$, and some (or all) of the $\mu_1(\cdot), ..., \mu_p(\cdot)$ might be nonlinear.

10. All the observed X_{2n} should be exogenous, independent of both $\boldsymbol{\varepsilon}_n$ and V_n. Refer to Park and Gupta (2012) for more detailed descriptions of this concern in empirical studies.

11. For nonparametric copula, the construction of an empirical copula (Deheuvels, 1979; Nelsen, 2006) by interpolation or approximation (Li et al., 1997)with selection criteria (e.g., Genest and Rivest, 1993; Barbe et al., 1996) can be considered, which are beyond the scope of this paper.

12. $G(v_i)$,$H_1(z_{1,i}), ..., H_p(z_{p,i})$ and $R_1(x_{1,i}), ..., R_k(x_{k,i})$ are uniform(0,1) random variables. $H_1(z_{1,i}), ..., H_p(z_{p,i})$ and $R_1(x_{1,i}), ..., R_k(x_{k,i})$ can be identified from the sample data, and $G(v_i)$ is assumed to be normal. The case of unknown $G(v_i)$ might be obtained by some approximations, for instance sieve approximation in Su and Jin (2018), however, due to its complexity, we leave it for future studies.

Sklar's theorem (Sklar, 1959), a flexible multivariate joint distribution can be constructed,

$$F(x_{1,i}, ..., x_{k,i}, z_{1,i}, ..., z_{p,i}, v_i) = C(U_{x_{1,i}}, ..., U_{x_{k,i}}, U_{z_{1,i}}, ..., U_{z_{p,i}}, U_{v_i}) \tag{7}$$

where $C(\cdot) : [0,1]^{p+k+1} \to [0,1]$ is a $p+k+1$-dimensional copula function, $U_{x_{1,i}} = R_1(x_{1,i}), ..., U_{x_{k,i}} = R_k(x_{k,i}), U_{z_{1,i}} = H_1(z_{1,i}), ..., U_{z_{p,i}} = H_p(z_{p,i}), U_{v_i} = G(v_i)$. From Eq.(7), the joint density function is

$$f(x_{1,i}, ..., x_{k,i}, z_{1,i}, ..., z_{p,i}, v_i) = c(U_{x_{1,i}}, ..., U_{x_{k,i}}, U_{z_{1,i}}, ..., U_{z_{p,i}}, U_{v_i}) g(v_i) \prod_{\iota=1}^{p} h_\iota(z_{\iota,i}) \prod_{\tau=1}^{k} r_\tau(x_{\tau,i}) \tag{8}$$

where $c(U_{x_{1,i}}, ..., U_{x_{k,i}}, U_{z_{1,i}}, ..., U_{z_{p,i}}, U_{v_i}) = \frac{\partial^{p+k+1} C(U_{x_{1,i}},...,U_{x_{k,i}},U_{z_{1,i}},...,U_{z_{p,i}},U_{v_i})}{\partial x_{1,i} \cdots \partial x_{k,i} \partial z_{1,i} \cdots \partial z_{p,i} \partial v_i}$, $g(\cdot)$, $h_\iota(\cdot)(\iota = 1, ..., p)$ and $r_\tau(\cdot)(\tau = 1, ..., k)$ are marginal density functions of $G(\cdot)$, $H_\iota(\cdot)(\iota = 1, ..., p)$ and $R_\tau(\cdot)(\tau = 1, ..., k)$ respectively. Although there are various copula functions in the statistics literature (Nelsen, 2006; Balakrishnan and Lai, 2009), as in many previous studies (Park and Gupta, 2012; Haschka, 2022, etc), we consider the Gaussian copula for detailed illustration.[13] The $(p + k + 1)$- dimensional Gaussian copula with a correlation matrix $P \in [-1,1]^{(k+p+1)\times(k+p+1)}$, where $P = \begin{pmatrix} P_x & P_{xz} & \mathbf{0}_{k\times 1} \\ P'_{xz} & P_z & \rho_{vz} \\ \mathbf{0}'_{k\times 1} & \rho'_{vz} & 1 \end{pmatrix}$ with $\rho_{vz} = (\rho_{vz_1,...,}\rho_{vz_p})'$ being a p dimensional vector, P_{xz} being a $k \times p$ matrix, P_z and P_x being $p \times p$ and $k \times k$ matrices respectively, can be written as

$$C_P^{Gaussian}(U_{x_{1,i}}, ..., U_{x_{k,i}}, U_{z_{1,i}}, ..., U_{z_{p,i}}, U_{v_i}) = \Phi_P(x^*_{1,i}, ..., x^*_{k,i}, z^*_{1,i}, ..., z^*_{p,i}, v^*_i) \tag{9}$$

where $x^*_{\tau,i} = \Phi^{-1}(U_{x_{\tau,i}})(\tau = 1, ..., k)$, $z^*_{\iota,i} = \Phi^{-1}(U_{z_{\iota,i}})(\iota = 1, ..., p)$, $v^*_i = \Phi^{-1}(U_{v_i})$, $\Phi^{-1}(\cdot)$ 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 Eq.(8) and (9), the joint density function of $x_i = (x_{1,i}, \ldots, x_{k,i})'$, $z_i = (z_{1,i}, \ldots, z_{p,i})'$ and v_i is

$$\begin{aligned} f(x'_i, z'_i, v_i) =& \frac{1}{|P|^{\frac{1}{2}}} \exp\left[-\frac{(x^*_{1,i}, ..., x^*_{k,i}, z^*_{1,i}, ..., z^*_{p,i}, v^*_i)(P^{-1} - I_{k+p+1})(x^*_{1,i}, ..., x^*_{k,i}, z^*_{1,i}, ..., z^*_{p,i}, v^*_i)'}{2} \right] \\ & \cdot g(v_i) \prod_{\iota=1}^{p} h_\iota(z_{\iota,i}) \prod_{\tau=1}^{k} r_\tau(x_{\tau,i}) \end{aligned} \tag{10}$$

13. As mentioned in Park and Gupta (2012), the Gaussian copula is general and robust for most applications (Song, 2000) and incorporates many desirable properties (Danaher and Smith, 2011), for instance with the empirical CDFs, the Gaussian copula model only depends on the rank-order if raw data and is invariant to strictly monotonic transformations of variables in x_i, z_i and v_i. We investigate some properties of the proposed Gaussian copula by Monte Carlo simulations in Section 6.

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

$$
\begin{aligned}
&\ln L_n(\{x_i', z_i', v_i\}|\lambda, \beta, \sigma_v^2, P) \\
&= -\frac{n}{2}\ln|P| - \frac{1}{2}\sum_{i=1}^{n}(x_{1,i}^*, ..., x_{k,i}^*, z_{1,i}^*, ..., z_{p,i}^*, v_i^*)(P^{-1} - I_{k+p+1})(x_{1,i}^*, ..., x_{k,i}^*, z_{1,i}^*, ..., z_{p,i}^*, v_i^*)' \\
&\quad + \sum_{i=1}^{n}\ln \phi_{(0,\sigma_v^2)}\left(v_i(\lambda, \beta)\right)
\end{aligned} \tag{11}
$$

where $\phi_{(0,\sigma_v^2)}(\cdot)$ is the normal density with mean 0 and variance σ_v^2. Note that the nonparametric densities $h_\iota\left(z_{\iota,i}\right)(\iota = 1, ..., p)$ and $r_\tau\left(x_{\tau,i}\right)(\tau = 1, ..., k)$ disappear from the log-likelihood function because they don't involve any parameters. For the endogenous heterogeneity specification (Eq.(4)), the scalar spatial coefficient λ should be replaced by the vector of parameters ζ. The pseudo-maximum likelihood estimation (PMLE) method based on Eq.(11) is provided in Section 3.

The Gaussian copula models $(x_{1,i}^*, ..., x_{k,i}^*, z_{1,i}^*, ..., z_{p,i}^*, v_i^*)'$ as the standard multivariate normal distribution with the correlation matrix P. When there is one exogenous variable $x_{1,i}^*$ and one endogenous variable $z_{1,i}^*$, $\begin{pmatrix} x_{1,i}^* \\ z_{1,i}^* \\ v_i^* \end{pmatrix} \sim \mathrm{N}\left(\mathbf{0}_{3\times 1}, \begin{bmatrix} 1 & \rho_{x_1z_1} & 0 \\ \rho_{x_1z_1} & 1 & \rho_{vz_1} \\ 0 & \rho_{vz_1} & 1 \end{bmatrix}\right)$, by Cholesky decomposition,

$$
\begin{pmatrix} x_{1,i}^* \\ z_{1,i}^* \\ v_i^* \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \rho_{x_1z_1} & \sqrt{1-\rho_{x_1z_1}^2} & 0 \\ 0 & \frac{\rho_{vz_1}}{\sqrt{1-\rho_{x_1z_1}^2}} & \sqrt{1-\frac{\rho_{vz_1}^2}{1-\rho_{x_1z_1}^2}} \end{pmatrix} \cdot \begin{pmatrix} \varpi_{1,i} \\ \varpi_{2,i} \\ \varpi_{3,i} \end{pmatrix}
$$

with $(\varpi_{1,i}, \varpi_{2,i}, \varpi_{3,i})' \sim N(\mathbf{0}_{3\times 1}, I_3)$, where I_3 is a 3×3 identity matrix. Given the joint distribution and $v_i = \Phi_{(0,\sigma_v^2)}^{-1}(\Phi(v_i^*)) = \sigma_v v_i^*$, where $\Phi_{(0,\sigma_v^2)}(\cdot)$ is the normal distribution with mean 0 and variance σ_v^2, we have

$$
z_{1,i}^* = \rho_{x_1z_1}x_{1,i}^* + \sqrt{1-\rho_{x_1z_1}^2}\varpi_{2,i} = x_{1,i}^*\tilde{\beta}_1 + u_i
$$

where $\tilde{\beta}_1 = \rho_{x_1z_1}$, $u_i = \sqrt{1-\rho_{x_1z_1}^2}\varpi_{2,i}$. Then Eq.(1) with endogeneous spatial weights can be rewritten as

$$
\begin{aligned}
y_i &= \lambda\sum_{j\neq i} w_{ij}y_j + x_{1,i}\beta_1 + \sigma_v v_i^* \\
&= \lambda\sum_{j\neq i} w_{ij}y_j + x_{1,i}\beta_1 + \frac{\sigma_v\rho_{vz_1}}{1-\rho_{x_1z_1}^2}(z_{1,i}^* - \rho_{x_1z_1}x_{1,i}^*) + \sigma_v\sqrt{1-\frac{\rho_{vz_1}^2}{1-\rho_{x_1z_1}^2}}\varpi_{3,i} \\
&= \lambda\sum_{j\neq i} w_{ij}y_j + x_{1,i}\beta_1 + \gamma_1 u_i + \sigma_v\sqrt{1-\frac{\rho_{vz_1}^2}{1-\rho_{x_1z_1}^2}}\varpi_{3,i}
\end{aligned}
$$

where $\gamma_1 = \frac{\sigma_v\rho_{vz_1}}{1-\rho_{x_1z_1}^2}$. As suggested in Yang et al. (2022), we can use u_i as an additional regressor to correct for the endogeneity bias[14], then W_n can be treated as predetermined or exogenous. The

14. Park and Gupta (2012) only account for correlation among $z_{1,i}^*$ and v_i, and $z_{1,i}^*$ is added rather than u_i, but as shown in Yang et al. (2022), $z_{1,i}^*$ can be correlated with $\varpi_{3,i}$.

above approach can be easily extended to accommodate high dimensional exogenous variables $x_i^* = (x_{1,i}^*, ..., x_{k,i}^*)'$ and endogenous variables $z_i^* = (z_{1,i}^*, ..., z_{p,i}^*)'$, i.e., $(x_i^{*\prime}, z_i^{*\prime}, v_i^*)' \sim N(\mathbf{0}_{(p+k+1)\times 1}, P)$. By Cholesky decomposition,

$$P = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \cdot \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix}' = \begin{pmatrix} L_{11}L_{11}' & * & * \\ L_{21}L_{11}' & L_{21}L_{21}' + L_{22}L_{22}' & * \\ L_{31}L_{11}' & L_{31}L_{21}' + L_{32}L_{22}' & L_{31}L_{31}' + L_{32}L_{32}' + L_{33}L_{33}' \end{pmatrix}$$

From this, we have that

$$\begin{pmatrix} x_i^* \\ z_i^* \\ v_i^* \end{pmatrix} = \begin{pmatrix} L_{11} = \text{Chol}(P_x) & \mathbf{0}_{k\times p} & \mathbf{0}_{k\times 1} \\ L_{21} = P_{xz}'(L_{11}')^{-1} & L_{22} = \text{Chol}(P_z - L_{21}L_{21}') & \mathbf{0}_{p\times 1} \\ L_{31} = \mathbf{0}_{k\times 1}' & L_{32} = \rho_{vz}'(L_{22}')^{-1} & L_{33} = \text{Chol}(1 - L_{32}L_{32}') \end{pmatrix} \cdot \begin{pmatrix} \varpi_{1,i} \\ \varpi_{2,i} \\ \varpi_{3,i} \end{pmatrix}$$

with $(\varpi_{1,i}, \varpi_{2,i}, \varpi_{3,i})' \sim N(\mathbf{0}_{(k+p+1)\times 1}, I_{k+p+1})$, where $\text{Chol}(\cdot)$ represents the Cholesky decomposition. Given the joint distribution, we have

$$z_i^* = L_{21}L_{11}^{-1}x_i^* + L_{22}\varpi_{2,i} = \Gamma' x_i^* + u_i \tag{12}$$

where $\Gamma = (P_{xz}' P_x^{-1})'$ and $u_i = L_{22}\varpi_{2,i}$. Then Eq.(1) can be rewritten as

$$\begin{aligned} y_i &= \lambda \sum_{j\neq i} w_{ij}y_j + x_i'\beta + \sigma_v v_i^* \\ &= \lambda \sum_{j\neq i} w_{ij}y_j + x_i'\beta + \sigma_v (z_i^* - \Gamma' x_i^*)'(L_{22}^{-1})' L_{32}' + \sigma_v L_{33}\varpi_{3,i} \\ &= \lambda \sum_{j\neq i} w_{ij}y_j + x_i'\beta + u_i'\gamma + \epsilon_i \end{aligned} \tag{13}$$

where $\gamma = \sigma_v (L_{22}^{-1})' L_{32}'$, $\epsilon_i = \eta\varpi_{3,i}$ with $\eta = \sigma_v L_{33}$.[15] Similarly, the specifications for Eq.(4) and (5) are

$$y_i = \lambda(\zeta, z_i) \sum_{j\neq i} w_{ij}y_j + x_i'\beta + u_i'\gamma + \epsilon_i \tag{14}$$

and

$$y_i = \lambda \sum_{j\neq i} w_{ij}y_j + x_i'\beta_1 + z_i\beta_2 + u_i'\gamma + \epsilon_i \tag{15}$$

respectively. We show in Section 3 that the same technique can be applied in the PMLE approach. By the above construction, we don't need to impose the implicit assumption that x_i and the pseudo-observations z_i^* are uncorrelated as we model the dependence structure between x_i^* and z_i^* directly. Moreover, the non-normality assumption has been relaxed for the third variant of a SAR model, z_i can be normally distributed and the model can be identified because u_i would not be perfectly collinear

15. This implied model differs from the partially linear SAR model (e.g., Su and Jin, 2010), i.e., $y_i = \lambda \sum_{j\neq i} w_{ij}y_j + x_i'\beta + m_0(z_i) + v_i$, where $m_0(\cdot)$ is an unknown function and the model can be estimated by a profile QMLE based on local polynomial procedure. While u_i are residuals by regressing z_i^* on x_i^*, where x_i^* and z_i^* are inverse (standard) normal distribution of unknown marginal distributions.

with z_i and x_i as long as one of the xs correlated with z_i is not normally distributed.[16] Eq.(13)-(15) indicate that the models can also be estimated by an IV estimation approach to account for the endogenous W_nY_n, the derivations are provided in the next section. To summarize, the constructed log pseudo-likelihood function (Eq.(11)) and the three equations (Eq.(13)-(15)) with u_i added to control for endogeneity impose the two assumptions below:

Assumption 1. *For each n, v_i's are i.i.d. $N(0, \sigma_v^2)$ random variables.*

Assumption 2. *(i) x_i, z_i, and v_i follow a Guassian copula, i.e., $(x_i^{*\prime}, z_i^{*\prime}, v_i^*)' \sim \mathrm{N}(\mathbf{0}_{(p+k+1)\times 1}, P)$. (ii) When endogenous z_i are added as regressors, one of the correlated exogenous regressors x_i should be non-normally distributed.*

3 Estimation

Because x_i and z_i are allowed to have high dimensions, and v_i is a scalar, it's more convenient to employ an equivalent version of Eq.(11) by shifting the orders of the three variables,

$$\begin{aligned}&\ln L_n(\{v_i, z_i', x_i', \}|\lambda, \beta, \sigma_v^2, \tilde{P}) \\ &= -\frac{n}{2}\ln|\tilde{P}| - \frac{1}{2}\sum_{i=1}^{n}(v_i^*, z_i^{*\prime}, x_i^{*\prime})(\tilde{P}^{-1} - I_{k+p+1})(v_i^*, z_i^{*\prime}, x_i^{*\prime})' + \sum_{i=1}^{n}\ln\phi_{(0,\sigma_v^2)}(v_i(\lambda, \beta))\end{aligned} \tag{16}$$

where the correlation matrix is redefined as $\tilde{P} = \begin{pmatrix} 1 & \rho_{vz}' & \mathbf{0}_{k\times 1}' \\ \rho_{vz} & P_z & P_{xz}' \\ \mathbf{0}_{k\times 1} & P_{xz} & P_x \end{pmatrix}$. By the partitioned quadratic formulation[17], we obtain

$$\begin{aligned}(v_i^*, z_i^{*\prime}, x_i^{*\prime})\tilde{P}^{-1}(v_i^*, z_i^{*\prime}, x_i^{*\prime})' &= \frac{1}{\kappa}\left[v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - P_{xz}'P_x^{-1}x_i^*)\right]'\left[v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - P_{xz}'P_x^{-1}x_i^*)\right] \\ &+ (z_i^* - P_{xz}'P_x^{-1}x_i^*)'\Xi^{-1}(z_i^* - P_{xz}'P_x^{-1}x_i^*) + x_i^{*\prime}P_x^{-1}x_i^*\end{aligned} \tag{17}$$

where $\Xi = P_z - P_{xz}'P_x^{-1}P_{xz}$, $\kappa = 1 - \rho_{vz}'\Xi^{-1}\rho_{vz}$. From Eq.(12), $\Gamma = (P_{xz}'P_x^{-1})'$, then

$$\begin{aligned}(v_i^*, z_i^{*\prime}, x_i^{*\prime})\tilde{P}^{-1}(v_i^*, z_i^{*\prime}, x_i^{*\prime})' &= \frac{1}{\kappa}\left[v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - \Gamma'x_i^*)\right]'\left[v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - \Gamma'x_i^*)\right] \\ &+ (z_i^* - \Gamma'x_i^*)'\Xi^{-1}(z_i^* - \Gamma'x_i^*) + x_i^{*\prime}P_x^{-1}x_i^*,\end{aligned} \tag{18}$$

16. Refer to the proof of Theorem 3 in Yang et al. (2022).
17. The derivation is provided in Appendix A.

Alternatively, the above log-likelihood function (Eq.(16)) can be written as

$$
\begin{aligned}
&\ln L_n\left(\{v_i, z_i', x_i'\} \mid \lambda, \beta, \sigma_v^2, \kappa, \rho_{vz}, P_x, \Xi\right) \\
&= -\frac{n}{2}\ln\kappa - \frac{n}{2}\ln|P_x| - \frac{n}{2}\ln|\Xi| - \frac{1}{2}\sum_{i=1}^{n}\Big[\kappa^{-1}\left(v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - \Gamma' x_i^*)\right)'\left(v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - \Gamma' x_i^*)\right) \\
&\quad + (z_i^* - \Gamma' x_i^*)'\Xi^{-1}(z_i^* - \Gamma' x_i^*) + x_i^{*\prime}(P_x^{-1} - I_k)x_i^* - v_i^{*2} - z_i^{*\prime} z_i^*\Big] + \sum_{i=1}^{n}\ln\phi_{(0,\sigma_v^2)}\left(v_i(\lambda,\beta)\right)
\end{aligned}
\tag{19}
$$

Refer to Appendix A.2 for the derivation of the determinant equality $|\tilde{P}| = |\kappa|\cdot|P_x|\cdot|\Xi|$. Since the transformed pseudo-observations x_i^*, z_i^* and the residuals $u_i = z_i^* - \Gamma' x_i^*$ in Eq.(13)-(15) and (19) are unobservables but can be consistently estimated, we propose a three-stage estimation method for the three model settings.

3.1 The first stage estimation

In the first stage, we get estimates for the marginal transformations $\overline{r}_\tau(x) = \Phi^{-1}(R_\tau(x))\,(\tau = 1, ..., k)$ and $\hbar_\iota(z) = \Phi^{-1}(H_\iota(z))\,(\iota = 1, ..., p)$. For the purpose of this paper, any estimation method that yields estimators $\hat{x}_{\tau,i}^*$ and $\hat{z}_{\iota,i}^*$ satisfying $\sup_{x_{\tau,i}}|\hat{x}_{\tau,i}^* - x_{\tau,i}^*| = o_p(1)$ and $\sup_{z_{\iota,i}}|\hat{z}_{\iota,i}^* - z_{\iota,i}^*| = o_p(1)$ can be chosen.[18] Let $\hat{R}_\tau(x) = \frac{1}{n}\sum_{i=1}^n I(x_{\tau,i} \le x)$ and $\hat{H}_\iota(z) = \frac{1}{n}\sum_{i=1}^n I(z_{\iota,i} \le z)$ be the empirical distribution functions of x_τ and z_ι. We consider the estimator proposed in Liu et al. (2012):

$$
\hat{\overline{r}}_\tau(x) := \Phi^{-1}\left(T_{1/(2n)}[\hat{R}_\tau(x)]\right) \text{ and } \hat{\hbar}_\iota(z) := \Phi^{-1}\left(T_{1/(2n)}[\hat{H}_\iota(z)]\right) \tag{20}
$$

where $T_{1/(2n)}[x] := \frac{1}{2n}\cdot I(x < \frac{1}{2n}) + x\cdot I(\frac{1}{2n} \le x \le 1 - \frac{1}{2n}) + (1 - \frac{1}{2n})\cdot I(x > 1 - \frac{1}{2n})$ 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, $\hat{x}_{\tau,i}^* = \hat{\overline{r}}_\tau(x_{\tau,i})(\tau = 1, ..., k)$ and $\hat{z}_{\iota,i}^* = \hat{\hbar}_\iota(z_{\iota,i})(\iota = 1, ..., p)$ for $i = 1, ..., n$.

3.2 The second stage estimation

Denote $X_n^* = (x_1^*, ..., x_n^*)'$, $Z_n^* = (z_1^*, ..., z_n^*)'$, and $U_n = (u_1, ..., u_n)'$ with $u_i = (u_{1,i}, ..., u_{p,i})'$, given $\hat{X}_n^*$ and $\hat{Z}_n^*$ from the first stage estimation, by substituting $\hat{X}_n^*$ for X_n^* and $\hat{Z}_n^*$ for Z_n^* in Eq.(12), we have

$$
\hat{Z}_n^* = \hat{X}_n^*\Gamma + \hat{U}_n \tag{21}
$$

where $\hat{U}_n = U_n - (Z_n^* - \hat{Z}_n^*) + (X_n^* - \hat{X}_n^*)\Gamma$. By the ordinary least squares (OLS) method, we can obtain the estimates for Γ,

$$
\hat{\Gamma} = (\hat{X}_n^{*\prime}\hat{X}_n^*)^{-1}\hat{X}_n^{*\prime}\hat{Z}_n^* \tag{22}
$$

18. Park and Gupta (2012) recommend using kernel density estimators to estimate all the marginal density functions, for example, $\hat{h}_\iota(z_\iota) = n^{-1}\sum_{i=1}^n K_{h_\iota}(z_\iota - z_{\iota,i})$, $\iota = 1, \ldots, p$, where $K_{h_\iota}(z_\iota) = h_\iota^{-1}K_\iota(z_\iota/h_\iota)$ with $K_\iota(\cdot)$ being a kernel function on $\mathbb{R}$ and $h_\iota = h_{\iota,n}$ being a bandwidth. Then $\hat{U}_{z_{\iota,i}} = \hat{H}_\iota(z_{\iota,i}) = \int_{-\infty}^{z_{\iota,i}}\hat{h}_\iota(z_\iota)\,dz_\iota$ and $\hat{z}_{\iota,i}^* = \Phi^{-1}\left(\hat{U}_{z_{\iota,i}}\right) = \sqrt{2}erf^{-1}(2\hat{U}_{z_{\iota,i}} - 1)$, where $erf(x) = 2\pi^{-\frac{1}{2}}\int_0^x e^{-t^2}dt$ is the error function. In order to achieve the uniform convergence requirement of $z_{\iota,i}^*$ (and $x_{\tau,i}^*$), a set of assumptions on the kernel functions, bandwidth sequences, and density functions $h_\iota(\cdot)$ should be imposed.

Then $\hat{Z}_n^* - \hat{X}_n^*\hat{\Gamma} = \hat{\mathscr{O}}_n^\perp \hat{Z}_n^* = \hat{\mathscr{O}}_n^\perp \hat{U}_n$, where $\hat{\mathscr{O}}_n^\perp = I_n - \hat{\mathscr{O}}_n$ with $\hat{\mathscr{O}}_n = \hat{X}_n^*(\hat{X}_n^{*\prime}\hat{X}_n^*)^{-1}\hat{X}_n^{*\prime}$.

3.3 The third stage estimation

3.3.1 3-Stage pseudo-maximum likelihood estimation

Denote $\omega = (\lambda, \beta')'$ (or $\omega = (\zeta', \beta')'$ for Eq.(4)), as $v_i^* = \Phi^{-1}\left(\Phi_{(0,\sigma_v^2)}(v_i(\omega))\right) = \frac{v_i(\omega)}{\sigma_v}$, given the estimates from the first and second stage estimations, note that

$$
\begin{aligned}
&-\frac{1}{2}\sum_{i=1}^n \left[\frac{1}{\kappa}\left(v_i^* - \rho_{vz}'\Xi^{-1}(\hat{z}_i^* - \hat{\Gamma}'\hat{x}_i^*)\right)'\left(v_i^* - \rho_{vz}'\Xi^{-1}(\hat{z}_i^* - \hat{\Gamma}'\hat{x}_i^*)\right) - v_i^{*2}\right] \\
&= -\frac{1}{2\kappa\sigma_v^2}\sum_{i=1}^n \left(v_i(\omega) - \sigma_v\rho_{vz}'\Xi^{-1}(\hat{z}_i^* - \hat{\Gamma}'\hat{x}_i^*)\right)'\left(v_i(\omega) - \sigma_v\rho_{vz}'\Xi^{-1}(\hat{z}_i^* - \hat{\Gamma}'\hat{x}_i^*)\right) + \frac{1}{2\sigma_v^2}\sum_{i=1}^n v_i(\omega)^2 \\
&= -\frac{1}{2\kappa\sigma_v^2}\left[V_n(\omega) - (\hat{\mathscr{O}}_n^\perp\hat{Z}_n^*)\varsigma\sigma_v\right]'\left[V_n(\omega) - (\hat{\mathscr{O}}_n^\perp\hat{Z}_n^*)\varsigma\sigma_v\right] + \frac{1}{2\sigma_v^2}V_n(\omega)'V_n(\omega),
\end{aligned}
$$

where $\varsigma = \Xi^{-1}\rho_{vz}$, and

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

where $V_n(\omega) = S_n(\lambda)Y_n - X_n\beta$. For Eq.(4) 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 $\Lambda(\zeta, Z_n) \equiv \mathrm{diag}\{\lambda(\zeta, z_1), ..., \lambda(\zeta, z_n)\}$. Therefore, the log-likelihood function (Eq.(19)) can be rewritten as

$$
\begin{aligned}
\ln L_n(\theta_{ML}) = &-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_\xi^2 - \frac{n}{2}\ln|P_x| - \frac{n}{2}\ln|\Xi| + \ln|S_n(\lambda)| - \frac{1}{2\sigma_\xi^2}\left[V_n(\omega) - (\hat{\mathscr{O}}_n^\perp\hat{Z}_n^*)\chi\right]'\left[V_n(\omega) - (\hat{\mathscr{O}}_n^\perp\hat{Z}_n^*)\chi\right] \\
&- \frac{1}{2}\sum_{i=1}^n(\hat{z}_i^* - \hat{\Gamma}'\hat{x}_i^*)'\Xi^{-1}(\hat{z}_i^* - \hat{\Gamma}'\hat{x}_i^*) - \frac{1}{2}\sum_{i=1}^n \hat{x}_i^{*\prime}(P_x^{-1} - I_k)\hat{x}_i^* + \frac{1}{2}\sum_{i=1}^n \hat{z}_i^{*\prime}\hat{z}_i^*
\end{aligned}
\tag{23}
$$

where $\theta_{ML} = (\omega', \sigma_\xi^2, \chi', \alpha', \delta')'$ with $\sigma_\xi^2 = \kappa\sigma_v^2$, $\chi = \sigma_v\varsigma$, α and δ being $J_1(=\frac{(k-1)(k-2)}{2})$ and J_2-dimensional column vectors of distinct elements in P_x and Ξ respectively. The 3SPML estimator $\hat{\theta}_{ML} = \arg\max\ln L_n(\theta_{ML})$.

Remark (Potential bias of 3SPML estimator for the third variant) In this case, $V_n(\omega) = S_n(\lambda)Y_n - X_n\beta_1 - Z_n\beta_2$, where $\omega = (\lambda, \beta_1', \beta_2')'$. From the first order derivatives[19] at $\theta_{ML,0}$ and the reduced form

19. Similar to those provided in Appendix A.3.2.

of Eq.(5), denote $G_n = W_n S_n^{-1}$, we have $\frac{\ln L_n(\theta_{ML,0})}{\partial \theta_{ML}} = \frac{\ln L_n^u(\theta_{ML,0})}{\partial \theta_{ML}} + \Lambda_n^{mle}$, where

$$\frac{\ln L_n^u(\theta_{ML,0})}{\partial \theta_{ML}} = \begin{pmatrix} \frac{1}{\sigma_{\xi,0}^2}\left[G_n(X_n\beta_{1,0}+V_n)\right]'\left[V_n - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi_0\right] - \text{tr}(G_n) \\ \frac{1}{\sigma_{\xi,0}^2}X_n'\left[V_n - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi_0\right] \\ \mathbf{0}_{p\times 1} \\ -\frac{n}{2\sigma_{\xi,0}^2} + \frac{1}{2\sigma_{\xi,0}^4}\left[V_n - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi_0\right]'\left[V_n - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi_0\right] \\ -\frac{n}{2}\frac{\partial \ln|P_{x,0}|}{\partial\alpha} - \frac{1}{2}\frac{\partial}{\partial\alpha}\text{tr}\left[P_{x,0}^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*\right] \\ -\frac{n}{2}\frac{\partial \ln|\Xi_0|}{\partial\delta} - \frac{1}{2}\frac{\partial}{\partial\delta}\text{tr}\left[\Xi_0^{-1}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\right] \end{pmatrix},$$

and

$$\Lambda_n^{mle} = \left(\frac{1}{\sigma_{\xi,0}^2}(G_nZ_n\beta_{2,0})'\left[V_n - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi_0\right], \mathbf{0}_{k\times 1}', \frac{1}{\sigma_{\xi,0}^2}Z_n'\left[V_n - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi_0\right], 0, \mathbf{0}_{J_1\times 1}', \mathbf{0}_{J_2\times 1}'\right)'.$$

Note that $\text{E}\left(\frac{1}{\sqrt{n}}\frac{\ln L_n^u(\theta_{ML,0})}{\partial\theta_{ML}}\right) = \mathbf{0}_{(k+p+J_1+J_2+2)\times 1}$, however, it's possible that $\text{E}(\frac{1}{\sqrt{n}}\Lambda_n^{mle}) \neq \mathbf{0}_{(k+p+J_1+J_2+2)\times 1}$ if $\text{E}\left(\frac{1}{\sqrt{n}}\frac{1}{\sigma_{\xi,0}^2}(G_nZ_n\beta_{2,0})'\left[V_n - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi_0\right]\right) \neq 0$ and $\text{E}\left(\frac{1}{\sqrt{n}}\frac{1}{\sigma_{\xi,0}^2}Z_n'\left[V_n - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi_0\right]\right) \neq \mathbf{0}_{p\times 1}$. To see this, suppose z_i is normally distributed, $z_i = \Phi_{\sigma_z}^{-1}(\Phi(z_i^*)) = \sigma_z z_i^*$, then $\text{E}\left(\frac{1}{\sqrt{n}}\frac{1}{\sigma_{\xi,0}^2}(G_nZ_n\beta_{2,0})'\left[V_n - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi_0\right]\right) = \frac{1}{\sqrt{n}}\frac{1}{\sigma_{\xi,0}^2}(\sigma_{v,0}-1)\sigma_{z,0}\beta_{2,0}'\rho_{vz,0}\text{tr}[\text{E}(G_n)]$ and $\text{E}\left(\frac{1}{\sqrt{n}}\frac{1}{\sigma_{\xi,0}^2}Z_n'\left[V_n - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi_0\right]\right) = \frac{\sqrt{n}}{\sigma_{\xi,0}^2}(\sigma_{v,0}-1)\sigma_{z,0}\rho_{vz,0}$. The two expecation components can deviate from zero unless $\sigma_{v,0} = 1$.

3.3.2 3-Stage IV estimation

Given the first and second stage estimations, for Eq.(13) and (14), we have

$$Y_n = \lambda W_nY_n + X_n\beta + (\hat{Z}_n^* - \hat{X}_n^*\hat{\Gamma})\gamma + \hat{\boldsymbol{\epsilon}}_n = \lambda W_nY_n + X_n\beta + \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\gamma + \hat{\boldsymbol{\epsilon}}_n \tag{24}$$

and

$$Y_n = \Lambda(\zeta, Z_n)W_nY_n + X_n\beta + (\hat{Z}_n^* - \hat{X}_n^*\hat{\Gamma})\gamma + \hat{\boldsymbol{\epsilon}}_n = \Lambda(\zeta, Z_n)W_nY_n + X_n\beta + \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\gamma + \hat{\boldsymbol{\epsilon}}_n \tag{25}$$

where $\hat{\boldsymbol{\epsilon}}_n = \boldsymbol{\epsilon}_n + \hat{\mathscr{O}}_nU_n\gamma + \hat{\mathscr{O}}_n^{\perp}(Z_n^* - \hat{Z}_n^*)\gamma - \hat{\mathscr{O}}_n^{\perp}(X_n^* - \hat{X}_n^*)\Gamma\gamma$. We consider the IV estimators for Eq.(24) and (25) separately.

For a SAR model with endogenous W_n (Eq.(24)), denote $\theta_{IV,w} = (\lambda, \beta', \gamma')'$, $\hat{M}_{n,w} = \left(W_nY_n, X_n, \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\right)$ and $\hat{T}_{n,w} = \left(Q_{n,w}, X_n, \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\right)$ [20], where $Q_{n,w}$ is an instrument variable matrix for the endogenous W_nY_n. As $W_nY_n = W_n(I_n - \lambda_0W_n)^{-1}(X_n\beta + U_n\gamma + \boldsymbol{\epsilon}_n)$ [21] and $(I_n - \lambda_0W_n)^{-1} = \sum_{l=0}^{\infty}\lambda_0^lW_n^l$, the column vectors of $Q_{n,w}$ can be linear combinations of $X_n, W_nX_n, W_n^2X_n, ...$ and columns in $\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*$.

20. $\theta_{IV} = (\lambda, \beta_1', \beta_2', \gamma')'$, $\hat{M}_n = \left(W_nY_n, X_n, Z_n, \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\right)$ and $T_n = \left(Q_n, X_n, Z_n, \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\right)$ for Eq.(15).
21. $W_nY_n = W_n(I_n - \lambda_0W_n)^{-1}(X_n\beta_1 + Z_n\beta_2 + U_n\gamma + \boldsymbol{\epsilon}_n)$ for the third variant.

The 3SIV estimator of $\theta_{IV,w}$ is

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

η_w^2 is a scalar and can be estimated by the sample average of the estimated residuals,

$$\hat{\eta}_w^2 = \frac{1}{n}(Y_n - \hat{\lambda}W_nY_n - X_n\hat{\beta} - \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\hat{\gamma})'(Y_n - \hat{\lambda}W_nY_n - X_n\hat{\beta} - \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\hat{\gamma}). \tag{27}$$

For a SAR model with endogenous heterogeneity (Eq.(25)), 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. [22] Denote $\theta_{IV,h} = (\zeta', \beta', \gamma')'$ with $\zeta = (\varrho_1, ..., \varrho_p)'$, $\hat{M}_{n,h} = \left(\Lambda_1(z_1)W_nY_n, ..., \Lambda_p(z_p)W_nY_n, X_n, \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\right)$ with $\Lambda_\iota(z_\iota) = \text{diag}\{F_\iota(z_{\iota,1}), ..., F_\iota(z_{\iota,n})\}(\iota = 1, ..., p)$, $\hat{T}_{n,h} = (Q_{1n,h}, ..., Q_{pn,h}, X_n, \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)$, where $Q_{\iota n,h}$ is an instrument variable matrix for $\Lambda_\iota(z_\iota)W_nY_n$. Because $\Lambda_\iota(z_\iota)W_nY_n = \Lambda_\iota(z_\iota)W_n[I_n - \sum_{\iota=1}^p \varrho_{\iota,0}\Lambda_\iota(z_\iota)W_n]^{-1} \cdot (X_n\beta + U_n\gamma + \boldsymbol{\epsilon}_n)$ and $[I_n - \sum_{\iota=1}^p \varrho_{\iota,0}\Lambda_\iota(z_\iota)W_n]^{-1} = \sum_{k=1}^{\infty}[\sum_{\iota=1}^p \varrho_{\iota,0}\Lambda_\iota(z_\iota)W_n]^k$, assuming the power series is well-defined, for which a sufficient condition is $\|\sum_{\iota=1}^p \varrho_{\iota,0}\Lambda_\iota(z_\iota)W_n\| < 1$, 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 $\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*$. We may also employ columns of X_n pre-multiplied by cross-products of the $\Lambda_\iota(z_\iota)W_n$ in view of the fact that $[\sum_{\iota=1}^p \varrho_{\iota,0}\Lambda_\iota(z_\iota)W_n]^k = \sum_{i_1+i_2+\cdots+i_p=k; i_1,i_2,...,i_p\geq 0}\begin{pmatrix} k \\ i_1, i_2, \ldots, i_p \end{pmatrix} \cdot \prod_{\iota=1}^p[\varrho_{\iota,0}\Lambda_\iota(z_\iota)W_n]^{i_\iota}$, where $\begin{pmatrix} k \\ i_1, i_2, \ldots, i_p \end{pmatrix} = \frac{k!}{i_1!i_2!\cdots i_p!}$ by the multinomial theorem. The 3SIV estimator of $\theta_{IV,h}$ is

$$\hat{\theta}_{IV,h} = \left[\hat{M}'_{n,h}\hat{T}_{n,h}\left(\hat{T}'_{n,h}\hat{T}_{n,h}\right)^{-1}\hat{T}'_{n,h}\hat{M}_{n,h}\right]^{-1}\hat{M}'_{n,h}\hat{T}_{n,h}\left(\hat{T}'_{n,h}\hat{T}_{n,h}\right)^{-1}\hat{T}'_{n,h}Y_n \tag{28}$$

and η_h^2 can be estimated by

$$\hat{\eta}_h^2 = \frac{1}{n}\left(Y_n - \sum_{\iota=1}^p \hat{\varrho}_\iota\Lambda_\iota(z_\iota)W_nY_n - X_n\hat{\beta} - \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\hat{\gamma}\right)'\left(Y_n - \sum_{\iota=1}^p \hat{\varrho}_\iota\Lambda_\iota(z_\iota)W_nY_n - X_n\hat{\beta} - \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\hat{\gamma}\right). \tag{29}$$

22. When $\lambda(\zeta, z_i) = \rho F(\varrho_0 + \varrho_1 z_{1,i} + \cdots + \varrho_p z_{p,i})$, the IV estimator would not be feasible because $\varrho_0, ..., \varrho_p$ in the nonlinear transformation $F(\cdot)$ are unknown.

4 Asymptotic analysis

4.1 The first stage estimator

The result that $\hat{\bar{r}}_\tau(x)$ and $\hat{\hbar}_\iota(z)$ converge respectively to $\bar{r}_\tau(x)$ and $\hbar_\iota(z)$ uniformly over some expanding intervals is established in Liu et al. (2012) (Theorem 4.6) as stated in the proposition below.

Proposition 1. *Let $\ell_\tau := \bar{r}_\tau^{-1}$ and $\ell_\iota := \hbar_\iota^{-1}$ be the inverse functions of $\bar{r}_\tau$ and $\hbar_\iota$.*[23] *For any $a_\tau, a_\iota \in (0,1)$, define $\Upsilon_n^\tau := [\ell_\tau(-\sqrt{\frac{7}{4}a_\tau \log n}), \ell_\tau(\sqrt{\frac{7}{4}a_\tau \log n})]$ and $\Upsilon_n^\iota := [\ell_\iota(-\sqrt{\frac{7}{4}a_\iota \log n}), \ell_\iota(\sqrt{\frac{7}{4}a_\iota \log n})]$. Then $\sup_{x\in\Upsilon_n^\tau}|\hat{\bar{r}}_\tau(x) - \bar{r}_\tau(x)| = O_p\left(\sqrt{\frac{\log\log n}{n^{1-a_\tau}}}\right) = o_p(1)$ and $\sup_{z\in\Upsilon_n^\iota}|\hat{\hbar}_\iota(z) - \hbar_\iota(z)| = O_p\left(\sqrt{\frac{\log\log n}{n^{1-a_\iota}}}\right) = o_p(1)$.*

In a follow-up paper (Han et al., 2013), they have a stronger result by extending the regions of Υ_n^τ and Υ_n^ι to be optimal, i.e., $\Upsilon_n^\tau := [\ell_\tau(-\sqrt{2a_\tau \log n}), \ell_\tau(\sqrt{2a_\tau \log n})]$ and $\Upsilon_n^\iota := [\ell_\iota(-\sqrt{2a_\iota \log n}), \ell_\iota(\sqrt{2a_\iota \log n})]$. As a result, we have $\sup_{x_{\tau,i}}\left|\hat{x}_{\tau,i}^* - x_{\tau,i}^*\right| = o_p(1)$ and $\sup_{z_{\iota,i}}\left|\hat{z}_{\iota,i}^* - z_{\iota,i}^*\right| = o_p(1)$, where we omit the notations for the expanding intervals for brevity.

4.2 The second stage estimator

In this subsection, we show that the sampling errors $\hat{x}_{\tau,i}^* - x_{\tau,i}^*$ and $\hat{z}_{\iota,i}^* - z_{\iota,i}^*$ are asymptotically negligible, i.e., the errors coming from the estimations of $x_{\tau,i}^*$ by $\hat{x}_{\tau,i}^*$ and $z_{\iota,i}^*$ by $\hat{z}_{\iota,i}^*$ are of order $o_p(1)$ in the second (and third) stage estimation. For any $m \times n$ matrix A, denote $\|A\| = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2\right)^{1/2}$ as the Frobenius norm. From the proof of Lemma A.4 in Liu et al. (2012), $\sup_x\|x_i^*\| \leq \wp_0(k)$, $\sup_x\|\hat{x}_i^*\| \leq \hat{\wp}_0(k)$, $\sup_z\|z_i^*\| \leq \wp_0(p)$, and $\sup_z\|\hat{z}_i^*\| \leq \hat{\wp}_0(p)$, where $\wp_0(k)$, $\hat{\wp}_0(k)$, $\wp_0(p)$, $\hat{\wp}_0(p)$ are sequences of constants satisfying $\wp_0^2(k)k/n \to 0$, $\hat{\wp}_0^2(k)k/n \to 0$, $\wp_0^2(p)p/n \to 0$ and $\hat{\wp}_0^2(p)p/n \to 0$ as $n \to \infty$, then we have one supporting lemma (Lemma 1) for the formal result in Proposition 2.

Lemma 1. *(i) $\frac{1}{n}\|X_n^*\|^2 \leq \tilde{C}_0\wp_0^2(k)$, $\frac{1}{n}\|\hat{X}_n^*\|^2 \leq \tilde{C}_0\hat{\wp}_0^2(k)$, $\frac{1}{n}\|Z_n^*\|^2 \leq \tilde{C}_0\wp_0^2(p)$, $\frac{1}{n}\|\hat{Z}_n^*\|^2 \leq \tilde{C}_0\hat{\wp}_0^2(p)$, where $\tilde{C}_0$ is a finite constant.*
(ii) $\frac{1}{n}\|\hat{X}_n^* - X_n^*\|^2 = \log\log n \cdot (\sum_{\tau=1}^k n^{a_\tau - 1}) = o_p(1)$, $\frac{1}{n}\|\hat{Z}_n^* - Z_n^*\|^2 = \log\log n \cdot (\sum_{\iota=1}^p n^{a_\iota - 1}) = o_p(1)$, $a_\tau, a_\iota \in (0,1)$ for $\tau = 1, ..., k$ and $\iota = 1, ..., p$.

Proposition 2. *$\frac{1}{n}\left[a'\varphi_n'(\theta)(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)b - a'\varphi_n'(\theta)(\mathscr{O}_n^\perp Z_n^*)b\right] = o_p(1)$, where a and b are conformable vectors of constants, $\varphi_n(\theta)$ is a bounded vector-value function or a conformable matrix, and $\mathscr{O}_n^\perp = I_n - X_n^*(X_n^{*\prime}X_n^*)^{-1}X_n^{*\prime}$.*

4.3 The third stage estimator

4.3.1 Assumptions and topological structure

To analyze the asymptotic properties of the 3SPML estimator and the 3SIV estimator, we need the following assumptions and topological specification.

23. Since $\bar{r}_\tau(x) = \Phi^{-1}(R_\tau(x))$ and $\hbar_\iota(z) = \Phi^{-1}(H_\iota(z))$, we have $\ell_\tau(\cdot) = R_\tau^{-1}(\Phi(\cdot))$ and $\ell_\iota(\cdot) = H_\iota^{-1}(\Phi(\cdot))$.

Assumption 3. *Individual units in the economy are located or living in a region* $D_n \subset D \subset \mathbb{R}^{d_0}$*, where the cardinality of* D_n *satisfies* $\lim_{n\to\infty}|D_n| = \infty$*. Any two different individual units* i *and* j *are located at distances of at least* $s_0 > 0$ *from each other, w.l.o.g. we assume that* $s_0 = 1$*.*

Assumption 4. *(i) For any* i, j *and* n*, the spatial weight* $w_{ij} \geq 0$*,* $w_{ii} = 0$*, and* $\sup_n \|W_n\|_\infty = c_w < \infty$*.*
(ii) Θ *denotes a compact parameter space for* θ*. The true parameter* θ_0 *is in the interior of* Θ*.*
(iii) For a SAR model with endogenous spatial weights, the spatial coefficient λ *satisfies* $\sup_{\lambda\in\Theta_\lambda}|\lambda| c_w < 1$*, where* Θ_λ *is the parameter space for* λ*. For a SAR model with endogenous heterogeneity,* $\lambda(\zeta, z)$ *is a globally bounded function with* $b_\lambda = \sup_{\zeta\in\Theta_\zeta, z}|\lambda(\zeta, z)| \leq \frac{1}{c_w}$ *and* $\lambda(\mathbf{0}, z) \equiv 0$ *for* $\forall z$*. Furthermore,* $\lambda(\zeta, z)$ *is smooth and strict monotonic for any element in* ζ *and* z *given all other elements are nonzero.*
(iv) V_n*,* X_n *and* Z_n *are stochastic and uniformly bounded,* $\sup_{n.i}\mathrm{E}|v_i|^{4+\mathfrak{d}} < \infty$*,* $\max_{\tau=1,\ldots,k}\sup_{n.i}\mathrm{E}|x_{\tau,i}|^{4+\mathfrak{d}} < \infty$*, and* $\max_{\iota=1,\ldots,p}\sup_{n.i}\mathrm{E}|z_{\iota,i}|^{4+\mathfrak{d}} < \infty$ *for some* $\mathfrak{d} > 0$*;* X_n *and* V_n *are independent.* $\lim_{n\to\infty}\frac{1}{n}X_n^{*\prime}X_n^*$ *exists and is nonsingular.*

Assumption 5. *The spatial weight* w_{ij} *satisfies* $0 \leq w_{ij} \leq c_1 d_{ij}^{-c_0 d_0}$ *for some* $c_1 \geq 0$ *and* $c_0 > 3 + \frac{1}{d_0}$*. Furthermore, there exist at most* $K (K \geq 1)$ *columns of* W_n *that the column sum exceeds* c_w*, where* K *is a fixed nonnegative integer that does not depend on* n*.*

Assumption 3 is the topological specification. The set D is a lattice of (possibly) unevenly placed locations, which can be a geographic (or economic, or their combined) space for a unit i. The distance may refer to physical and/or economic distance. The minimum distance s_0 is used to avoid extreme influence between two units and indicates that our asymptotic analysis is based on inference under spatial near-epoch dependence (NED) using increasing domain rather than infill domain asymptotics. Assumptions 4(*i*) and 4(*ii*) for the endogenous W_n specification are standard assumptions in the spatial econometrics literature, which are imposed to limit the spatial correlation in a manageable degree. Assumption 4(*iii*) for the endogenous heterogeneity case assumes that $\lambda(\zeta, z)$ is bounded, which restricts the parameter space Θ_ζ of ζ and the $\lambda(\zeta, z)$ function that can be chosen, and that the derivative functions of $\lambda(\zeta, z)$ exist and are bounded. Similar assumption is also imposed in Xu and Lee (2015a), they state a set of functions that will satisfy this assumption. Assumption 4(*iv*) gives regularity conditions for V_n, Z_n and X_n. The assumption imposed on $\lim_{n\to\infty}\frac{1}{n}X_n^{*\prime}X_n^*$ is required for the consistency of the second stage OLS estimator. Assumption 5 allows units far apart to correlate with each other, but the spatial weight should decrease sufficiently fast at a certain rate as the distance d_{ij} increases. This assumption accommodates a special case that $w_{ii} = 0$ if d_{ij} exceeds some threshold. The second part of this assumption puts limitation on the number of columns which can have larger magnitudes than the row sum norm, which allows the existence of some "stars", i.e., larger units that have larger aggregated impacts on other units even when n increases.

4.3.2 Asymptotic properties of the 3SPML estimator

Denote $\ln L_{n0}(\theta_{ML}) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_\xi^2 - \frac{n}{2}\ln|P_x| - \frac{n}{2}\ln|\Xi| + \ln|S_n(\lambda)| - \frac{1}{2\sigma_\xi^2}[V_n(\omega) - (\mathscr{O}_n^\perp Z_n^*)\chi]'[V_n(\omega) - (\mathscr{O}_n^\perp Z_n^*)\chi] - \frac{1}{2}\sum_{i=1}^n (z_i^* - \Gamma' x_i^*)'\Xi^{-1}(z_i^* - \Gamma' x_i^*) - \frac{1}{2}\sum_{i=1}^n x_i^{*\prime}(P_x^{-1} - I_k)x_i^* + \frac{1}{2}\sum_{i=1}^n z_i^{*\prime} z_i^*$, where the estimates from the first and second stage estimations are evaluated at their true values. In Lemma 2, we prove that the main difference between $\ln L_{n0}(\theta_{ML})$ and $\ln L_n(\theta_{ML})$ comes from terms related to $\hat{x}_i^*$ versus x_i^*, $\hat{z}_i^*$ versus z_i^*, and $\hat{\Gamma}$ versus Γ (as well as the differences in their sample averages of the first and second order derivatives[24]) are all $o_p(1)$.

Lemma 2. *Under Assumption 1, 2, 4(ii), and 4(iv),* $\frac{1}{n}\left[\ln L_n(\theta_{ML}) - \ln L_{n0}(\theta_{ML})\right] = o_p(1)$, $\frac{1}{n}\left[\frac{\partial \ln L_n(\theta_{ML})}{\partial \theta_l} - \frac{\partial \ln L_{n0}(\theta_{ML})}{\partial \theta_l}\right] = o_p(1)$, *and* $\frac{1}{n}\left[\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \theta_{l_1}\partial \theta_{l_2}} - \frac{\partial^2 \ln L_{n0}(\theta_{ML})}{\partial \theta_{l_1}\partial \theta_{l_2}}\right] = o_p(1)$, *where* $\theta_l, \theta_{l_1}, \theta_{l_2} = \lambda, \beta, \sigma_\xi^2, \chi, \alpha, \delta$ *for the endogenous spatial weights specification, and* $\theta_l, \theta_{l_1}, \theta_{l_2} = \zeta_1, \cdots, \zeta_{p_0}, \beta, \sigma_\xi^2, \chi, \alpha, \delta$ *for the endogenous heterogeneity specification.*

Denote $\boldsymbol{X}_n = \left(X_n, \left(\mathscr{O}_n^\perp Z_n^*\right)\right)$, Assumption 6 below is an identification condition for the model. Lemma 3 is a sufficient identification result (Rothenberg, 1971).

Assumption 6. *(i)* $\boldsymbol{X}_n'\boldsymbol{X}_n$ *is invertible with probability one.*
(ii) $S_n(\lambda)'S_n(\lambda)$ *is not proportional to* $S_n'S_n$ *with probability one whenever* $\lambda \neq \lambda_0$ *for a SAR model with endogenous spatial weights*[25], *where* $S_n(\lambda) = I_n - \lambda W_n$ *and* $S_n = S_n(\lambda_0)$.
(iii) $S_n(\zeta)'S_n(\zeta)$ *is not proportional to* $S_n'S_n$ *with probability one whenever* $\zeta \neq \zeta_0$ *for a SAR model with endogenous heterogeneity, where* $S_n(\zeta) = I_n - \Lambda(\zeta, Z_n)W_n$ *and* $S_n = S_n(\zeta_0)$.

Lemma 3. *Under Assumption 1, 2, 4(ii)-4(iii), and 6,* $\theta_{ML,0}$ *is the unique maximizer of* $\lim_{n\to\infty}\frac{1}{n}\ln L_n(\theta_{ML})$.

Assumption 7 is imposed to strengthen the identification information inequality to the limit. Assumption 8 is a regularity condition. Then by using the LLN (Theorem 1) and CLT (Theorem 2 and Corollary 1) in Jenish and Prucha (2012), we show the asymptotic properties of the 3SPML estimator in Theorem 1.

Assumption 7. $\limsup_{n\to\infty}\left[\mathrm{E}\ln L_n(\theta_{ML}) - \mathrm{E}\ln L_n(\theta_{ML,0})\right] < 0$ *for any* $\theta_{ML} \neq \theta_{ML,0}$.

Assumption 8. *(i)* $\{v_i, z_i, x_i\}_{i=1}^n$ *is an* α*-mixing random field with* α*-mixing coefficient* $\alpha(\mathfrak{p}, \mathfrak{q}, \mathfrak{r}) \leq (\mathfrak{p}+\mathfrak{q})^{\mathfrak{d}}\hat{\alpha}(\mathfrak{r})$ *for some* $\mathfrak{d} \geq 0$, *where* $\hat{\alpha}(\mathfrak{r})$ *satisfies* $\sum_{\mathfrak{r}=1}^\infty \mathfrak{r}^{d_0-1}\hat{\alpha}(\mathfrak{r}) < \infty$.
(ii) For some $\mathfrak{t} > 0$, *the* α*-mixing coefficient of* $\{v_i, z_i, x_i\}_{i=1}^n$ *satisfies* $\sum_{\mathfrak{r}=1}^\infty \mathfrak{r}^{d_0(\mathfrak{d}_*+1)}\hat{\alpha}^{\frac{\mathfrak{t}}{4+2\mathfrak{t}}}(\mathfrak{r}) < \infty$, *where* $\mathfrak{r} = \mathfrak{t}\mathfrak{d}/(2+\mathfrak{t})$.
(iii) $\lim_{n\to\infty}\frac{1}{n}\mathrm{E}\left(\frac{\partial^2 \ln L_n(\theta_{ML,0})}{\partial\theta_{ML}\partial\theta_{ML}'}\right)$ *exists and is nonsingular.*

Theorem 1. *Under Assumptions 1-8, the 3SPML estimator* $\hat{\theta}_{ML}$ *is a consistent estimator of* $\theta_{ML,0}$ *and* $\sqrt{n}(\hat{\theta}_{ML} - \theta_{ML,0}) \xrightarrow{d} \mathrm{N}(0, \mathcal{I}_{ML,0}^{-1})$, *where* $\mathcal{I}_{ML,0}^{-1} = -\left(\lim_{n\to\infty}\frac{1}{n}\mathrm{E}\left(\frac{\partial^2 \ln L_n(\theta_{ML,0})}{\partial\theta_{ML}\partial\theta_{ML}'}\right)\right)^{-1}$.

4.3.3 Asymptotic properties of the 3SIV estimator

Assumption 9. *Let* $M_{n,w} = (W_nY_n, X_n, \mathscr{O}_n^\perp Z_n^*)$, $T_{n,w} = (Q_{n,w}, X_n, \mathscr{O}_n^\perp Z_n^*)$, $M_{n,h} = (\Lambda_1(z_1)W_nY_n, ..., \Lambda_p(z_p)W_nY_n, X_n, \mathscr{O}_n^\perp Z_n^*)$, *and* $T_{n,h} = (Q_{1n,h}, ..., Q_{pn,h}, X_n, \mathscr{O}_n^\perp Z_n^*)$, *where* $Q_{n,w}$ *is an instrument variable matrix for* W_nY_n, $\Lambda_\iota(z_\iota) = \mathrm{diag}\{F_\iota(z_{\iota,1}), \ldots, F_\iota(z_{\iota,n})\}$ $(\iota = 1, \ldots, p)$ *and* $Q_{\iota n,h}$ *is an instrument variable matrix for* $\Lambda_\iota(z_\iota)W_nY_n$. $\mathrm{plim}_{n\to\infty}\frac{1}{n}[M_{n,w}'T_{n,w}(T_{n,w}'T_{n,w})^{-1}T_{n,w}'M_{n,w}]$ *and* $\mathrm{plim}_{n\to\infty}\frac{1}{n}[M_{n,h}'T_{n,h}(T_{n,h}'T_{n,h})^{-1}T_{n,h}'M_{n,h}]$ *exist and are nonsingular.*

24. The first and second order derivatives of $\ln L_n(\theta_{ML})$ can be found in Appendix A.3.2.
25. A sufficient condition is that matrices I_n, $(W_n + W_n')$ and $W_n'W_n$ are linearly independent. Refer to the proof in footnote 9 in Qu and Lee (2015).

Assumption 9 requires that all regressors, including the additional regressor $\mathscr{O}_n^{\perp} Z_n^*$ are asymptotically linearly independent such that the IV estimation in the third stage will not suffer from the collinearity problem. Then we have the asymptotic properties for the 3SIV estimator in Theorem 2.

Theorem 2. *Under Assumptions 1-5, 8(i)-8(ii) and 9, the 3SIV estimator $\hat{\theta}_{IV,\cdot}$ is a consistent estimator of $\theta_{IV,\cdot,0}$ and $\sqrt{n}(\hat{\theta}_{IV,\cdot} - \theta_{IV,\cdot,0}) \xrightarrow{d} \mathrm{N}(0, \mathcal{U}_{IV,\cdot})$, where*

$$\mathcal{U}_{IV,\cdot} = \text{plim}_{n\to\infty}\frac{1}{n}[M'_{n,\cdot}T_{n,\cdot}(T'_{n,\cdot}T_{n,\cdot})^{-1}T'_{n,\cdot}M_{n,\cdot}]^{-1}M'_{n,\cdot}T_{n,\cdot}(T'_{n,\cdot}T_{n,\cdot})^{-1}T'_{n,\cdot}\mathscr{G}_{n,\cdot}T_{n,\cdot}(T'_{n,\cdot}T_{n,\cdot})^{-1}T'_{n,\cdot}M_{n,\cdot} \cdot [M'_{n,\cdot}T_{n,\cdot}(T'_{n,\cdot}T_{n,\cdot})^{-1}T'_{n,\cdot}M_{n,\cdot}]^{-1},$$

$\cdot$ can be w and h, which represent endogenous spatial weights setting and endogenous heterogeneity specification respectively, and the variance of the composite error $\hat{\boldsymbol{\epsilon}}_n$ is $\mathscr{G}_{n,\cdot} = \eta^2_{\cdot,0}I_n + \gamma'_{\cdot,0}\Xi_{\cdot,0}\gamma_{\cdot,0}\hat{\mathscr{O}}_n$, where $\Xi_0 = P_{z,0} - P'_{xz,0}P^{-1}_{x,0}P_{xz,0}$.

As $\eta^2_{\cdot}$ is not directly estimated in the IV approach, the result below shows its consistent estimator (as constructed in Eq.(27) and (29)) and the consistently estimated asymptotic variance of $\hat{\theta}_{IV,\cdot}$.

Lemma 4. *Suppose $\hat{\theta}_{IV,\cdot}$ is a consistent estimator of $\theta_{IV,\cdot,0}$, then $\hat{\eta}^2_{\cdot} = \frac{1}{n}\tilde{\boldsymbol{\epsilon}}'_{n,\cdot}\tilde{\boldsymbol{\epsilon}}_{n,\cdot}$ is a consistent estimator of $\eta^2_{\cdot,0}$, where $\tilde{\boldsymbol{\epsilon}}_{n,w} = Y_n - \hat{\lambda}W_nY_n - X_n\hat{\beta} - \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\hat{\gamma}$ and $\tilde{\boldsymbol{\epsilon}}_{n,h} = Y_n - \sum_{\iota=1}^{p}\hat{\varrho}_\iota\Lambda_\iota(z_\iota)W_nY_n - X_n\hat{\beta} - \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*\hat{\gamma}$. If $\theta_{IV,\cdot,0}$ is replaced with $\hat{\theta}_{IV,\cdot,0}$ in $\hat{M}_{n,\cdot}$ and $\mathscr{O}_n^{\perp}Z_n^*$ with $\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*$, the consistently estimated asymptotic variance-covariance matrices for the 3SIV estimator $\hat{\theta}_{IV,\cdot}$ is*

$$\frac{1}{n}\hat{\mathcal{U}}_{IV,\cdot} = \left[\hat{M}'_{n,\cdot}\hat{T}_{n,\cdot}\left(\hat{T}'_{n,\cdot}\hat{T}_{n,\cdot}\right)^{-1}\hat{T}'_{n,\cdot}\hat{M}_{n,\cdot}\right]^{-1}\hat{M}'_{n,\cdot}\hat{T}_{n,\cdot}\left(\hat{T}'_{n,\cdot}\hat{T}_{n,\cdot}\right)^{-1}\hat{T}'_{n,\cdot}\hat{\mathscr{G}}_{n,\cdot} \cdot \hat{T}_{n,\cdot}\left(\hat{T}'_{n,\cdot}\hat{T}_{n,\cdot}\right)^{-1}\hat{T}'_{n,\cdot}\hat{M}_{n,\cdot}\left[\hat{M}'_{n,\cdot}\hat{T}_{n,\cdot}\left(\hat{T}'_{n,\cdot}\hat{T}_{n,\cdot}\right)^{-1}\hat{T}'_{n,\cdot}\hat{M}_{n,\cdot}\right]^{-1},$$

where $\hat{\mathscr{G}}_{n,\cdot} = \hat{\eta}^2_{\cdot,0}I_n + \hat{\gamma}'_{\cdot,0}\hat{\Xi}_{\cdot,0}\hat{\gamma}_{\cdot,0}\hat{\mathscr{O}}_n$ with $\hat{\Xi}_{\cdot,0} = \frac{1}{n}\hat{Z}_n^{*\prime}\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*$.

5 Extension

In this section, we extend the endogenous heterogeneity specification to the case with group fixed effects.[26] One application of this model would be the spatial spillovers of opioid sales, for instance Guth and Zhang (2022) investigate geographic spillover effects of prescription drug monitoring programs (PDMPs) (although the spatial interaction term is not included directly), it might be meaningful to account for state-level fixed effects and get comparable estimates (versus those obtained using the second variant of a SAR model) as the PDMPs vary across states.

Suppose the n individual units belong to G groups $\{\mathcal{G}_n^g\}_{g=1}^G$ which satisfy $\cup_{g=1}^G \mathcal{G}_n^g = n$ and $\mathcal{G}_n^{g_1} \cap$

26. The derivation for the first variant of a SAR model with group fixed effects follow a similar fashion. We don't consider the third variant because it's less interesting and there might exist asymptotic bias of the 3SPML estimator for this specification as mentioned in the remark of Section 3.3.1.

$\mathcal{G}_n^{g_2} = 0$ for $g_1 \neq g_2$. We consider the following setting,

$$y_i = \lambda(\zeta, z_i) \sum_{j \neq i} w_{ij} y_j + x_i' \beta + \sum_{g=1}^{G} r_{i,g} \boldsymbol{c}_g + v_i \tag{30}$$

where $r_{i,g}$ is an indicator for whether individual unit i belongs to group $\mathcal{G}_n^g$ or not, i.e., $r_{i,g} = 1$ if $i \in \mathcal{G}_n^g$ and $r_{i,g} = 0$ if $i \notin \mathcal{G}_n^g$; $\boldsymbol{c}_g$ denotes group fixed effect for group $\mathcal{G}_n^g$. We consider the 3SPML estimation, based on Eq.(23), the log-likelihood function is

$$\begin{aligned}\ln L_n(\theta_{ML}, \boldsymbol{c}) = &-\frac{n}{2}\ln(2\pi\sigma_\xi^2) - \frac{n}{2}\ln|P_x| - \frac{n}{2}\ln|\Xi| + \ln|S_n(\zeta)| - \frac{1}{2\sigma_\xi^2}\left[V_n(\omega, \boldsymbol{c}) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right]'\left[V_n(\omega, \boldsymbol{c}) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right] \\ &- \frac{1}{2}\sum_{i=1}^{n}(\hat{z}_i^* - \hat{\Gamma}'\hat{x}_i^*)'\Xi^{-1}(\hat{z}_i^* - \hat{\Gamma}'\hat{x}_i^*) - \frac{1}{2}\sum_{i=1}^{n}\hat{x}_i^{*\prime}(P_x^{-1} - I_k)\hat{x}_i^* + \frac{1}{2}\sum_{i=1}^{n}\hat{z}_i^{*\prime}\hat{z}_i^*\end{aligned} \tag{31}$$

where $\theta_{ML} = (\omega', \sigma_\xi^2, \chi', \alpha', \delta')'$ with $\omega = (\zeta', \beta')'$ and $\boldsymbol{c} = (\boldsymbol{c}_1, ..., \boldsymbol{c}_G)'$, $V_n(\omega, \boldsymbol{c}) = V_n(\omega) - R_G\boldsymbol{c}$ with $V_n(\omega) = S_n(\zeta)Y_n - X_n\beta$, $S_n(\zeta) = I_n - \Lambda(\zeta, Z_n)W_n$, and $R_G = (r_1, ..., r_G)'$ with $r_g = (r_{1,g}, ..., r_{n,g})'$. By the first-order conditions with respect to $\boldsymbol{c}$, we obtain

$$\hat{\boldsymbol{c}} = (R_G'R_G)^{-1}R_G'\left(V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right)$$

Hence,

$$V_n(\omega, \boldsymbol{c}) - \hat{Z}_n^*\varphi = \tilde{R}_G\left(V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right)$$

with $\tilde{R}_G = I_n - R_G(R_G'R_G)^{-1}R_G'$, where $R_G'R_G = \mathrm{diag}(m_1, ..., m_G)$, and m_g is the number of individuals units in group $\mathcal{G}_n^g$ for $g = 1, ..., G$. Then, the concentrated log-likelihood function for estimating θ_{ML} is

$$\begin{aligned}\ln L_n(\theta_{ML}) = &-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_\xi^2 - \frac{n}{2}\ln|P_x| - \frac{n}{2}\ln|\Xi| + \ln|S_n(\zeta)| \\ &- \frac{1}{2\sigma_\xi^2}\left[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right]'(\tilde{R}_G'\tilde{R}_G)\left[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right] \\ &- \frac{1}{2}\sum_{i=1}^{n}(\hat{z}_i^* - \hat{\Gamma}'\hat{x}_i^*)'\Xi^{-1}(\hat{z}_i^* - \hat{\Gamma}'\hat{x}_i^*) - \frac{1}{2}\sum_{i=1}^{n}\hat{x}_i^{*\prime}(P_x^{-1} - I_k)\hat{x}_i^* + \frac{1}{2}\sum_{i=1}^{n}\hat{z}_i^{*\prime}\hat{z}_i^*\end{aligned} \tag{32}$$

The first-order conditions at $\theta_{ML,0}$ are

$$\frac{\partial \ln L_n(\theta_{ML,0})}{\partial \theta_{ML}} = \begin{pmatrix} \frac{\partial \ln L_n(\theta_{ML,0})}{\partial \zeta_1} \\ \vdots \\ \frac{\partial \ln L_n(\theta_{ML,0})}{\partial \zeta_{p_0}} \\ \frac{\partial \ln L_n(\theta_{ML,0})}{\partial \beta} \\ \frac{\partial \ln L_n(\theta_{ML,0})}{\partial \sigma_\xi^2} \\ \frac{\partial \ln L_n(\theta_{ML,0})}{\partial \chi} \\ \frac{\partial \ln L_n(\theta_{ML,0})}{\partial \alpha} \\ \frac{\partial \ln L_n(\theta_{ML,0})}{\partial \delta} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma_{\xi,0}^2}[\mathrm{diag}(\frac{\partial \lambda(\zeta_0,z_i)}{\partial \zeta_1})W_nY_n]'(\tilde{R}_G'\tilde{R}_G)(V_n - \hat{Z}_n^*\varphi_0) - \mathrm{tr}[\mathrm{diag}(\frac{\partial \lambda(\zeta_0,z_i)}{\partial \zeta_1})W_nS_n^{-1}] \\ \vdots \\ \frac{1}{\sigma_{\xi,0}^2}[\mathrm{diag}(\frac{\partial \lambda(\zeta_0,z_i)}{\partial \zeta_{p_0}})W_nY_n]'(\tilde{R}_G'\tilde{R}_G)(V_n - \hat{Z}_n^*\varphi_0) - \mathrm{tr}[\mathrm{diag}(\frac{\partial \lambda(\zeta_0,z_i)}{\partial \zeta_{p_0}})W_nS_n^{-1}] \\ \frac{1}{\sigma_{\xi,0}^2}X_n'(\tilde{R}_G'\tilde{R}_G)(V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0) \\ -\frac{n}{2\sigma_{\xi,0}^2} + \frac{1}{2\sigma_{\xi,0}^4}(V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi)'(\tilde{R}_G'\tilde{R}_G)(V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0) \\ \frac{1}{\sigma_{\xi,0}^2}(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)^{*\prime}(\tilde{R}_G'\tilde{R}_G)(V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0) \\ -\frac{n}{2}\frac{\partial \ln|P_{x,0}|}{\partial \alpha} - \frac{1}{2}\frac{\partial}{\partial \alpha}\mathrm{tr}[P_{x,0}^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*] \\ -\frac{n}{2}\frac{\partial \ln|\Xi|}{\partial \delta} - \frac{1}{2}\frac{\partial}{\partial \delta}\mathrm{tr}[\Xi^{-1}(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)'(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)] \end{pmatrix}. \tag{33}$$

From items in (33) and the reduced form of Eq.(30), $Y_n = (I_n - \Lambda(\zeta_0, Z_n)W_n)^{-1}(X_n\beta_0 + R_G\boldsymbol{c}_0 + V_n)$, denote $\tilde{G}_{n,\iota} = \mathrm{diag}(\frac{\partial \lambda(\zeta_0,z_i)}{\partial \zeta_\iota})G_n$ with $G_n = W_nS_n^{-1}$, we have $\frac{\partial \ln L_n(\theta_{ML,0})}{\partial \theta_{ML}} = \frac{\partial \ln L_n^\nu(\theta_{ML,0})}{\partial \theta_{ML}} + \Delta_n$, where

$$\frac{\partial \ln L_n^\nu(\theta_{ML,0})}{\partial \theta_{ML}} = \begin{pmatrix} \frac{1}{\sigma_{\xi,0}^2}[\tilde{G}_{n,1}(X_n\beta_0 + R_G\boldsymbol{c}_0 + V_n)]'(\tilde{R}_G'\tilde{R}_G)(V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0) - \mathrm{tr}[(\tilde{R}_G'\tilde{R}_G)\tilde{G}_{n,1}] \\ \vdots \\ \frac{1}{\sigma_{\xi,0}^2}[\tilde{G}_{n,p_0}(X_n\beta_0 + R_G\boldsymbol{c}_0 + V_n)]'(\tilde{R}_G'\tilde{R}_G)(V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0) - \mathrm{tr}[(\tilde{R}_G'\tilde{R}_G)\tilde{G}_{n,p_0}] \\ \frac{1}{\sigma_{\xi,0}^2}X_n'(\tilde{R}_G'\tilde{R}_G)(V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0) \\ \frac{1}{2\sigma_{\xi,0}^4}(V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0)'(\tilde{R}_G'\tilde{R}_G)(V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0) - \frac{1}{2\sigma_{\xi,0}^2}\mathrm{tr}(\tilde{R}_G'\tilde{R}_G) \\ \frac{1}{\sigma_{\xi,0}^2}(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)^{*\prime}(\tilde{R}_G'\tilde{R}_G)(V_n - (\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)\chi_0) \\ -\frac{n}{2}\frac{\partial \ln|P_{x,0}|}{\partial \alpha} - \frac{1}{2}\frac{\partial}{\partial \alpha}\mathrm{tr}[P_{x,0}^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*] \\ -\frac{n}{2}\frac{\partial \ln|\Xi|}{\partial \delta} - \frac{1}{2}\frac{\partial}{\partial \delta}\mathrm{tr}[\Xi^{-1}(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)'(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)] \end{pmatrix}, \tag{34}$$

and

$$\Delta_n = \left(-\mathrm{tr}\left[(I_n - \tilde{R}_G'\tilde{R}_G)\tilde{G}_{n,1}\right], \cdots, -\mathrm{tr}\left[(I_n - \tilde{R}_G'\tilde{R}_G)\tilde{G}_{n,p_0}\right], \mathbf{0}_{k\times 1}', -\frac{1}{2\sigma_{\xi,0}^2}\mathrm{tr}(I_n - \tilde{R}_G'\tilde{R}_G), \mathbf{0}_{p\times 1}', \mathbf{0}_{J_1\times 1}', \mathbf{0}_{J_2\times 1}'\right)'. \tag{35}$$

The component $\frac{\partial \ln L_n^\nu(\theta_{ML,0})}{\partial \theta_{ML}} = \frac{\partial \ln L_n(\theta_{ML,0})}{\partial \theta_{ML}} - \mathrm{E}(\frac{\partial \ln L_n(\theta_{ML,0})}{\partial \theta_{ML}})$ has zero mean and is asymptotic normally distributed, $\Delta_n = \mathrm{E}(\frac{\partial \ln L_n(\theta_{ML,0})}{\partial \theta_{ML}})$ is the mean part that may cause asymptotic bias of the 3SPML estimator. Note that $\Delta_n = O(1)$. As $\tilde{R}_G'\tilde{R}_G = \tilde{R}_G$ and $\mathrm{tr}(\tilde{R}_G) = n - G$, define

$$\Delta_n(\theta_{ML}) = \left(-\mathrm{tr}\left[(I_n - \tilde{R}_G'\tilde{R}_G)\tilde{G}_{n,1}(\theta_{ML})\right], \cdots, -\mathrm{tr}\left[(I_n - \tilde{R}_G'\tilde{R}_G)\tilde{G}_{n,p_0}(\theta_{ML})\right], \mathbf{0}_{k\times 1}', -\frac{G}{2\sigma_{\xi,0}^2}, \mathbf{0}_{p\times 1}', \mathbf{0}_{J_1\times 1}', \mathbf{0}_{J_2\times 1}'\right)' \tag{36}$$

Then, the asymptotic distribution of $\hat{\theta}_{ML}$ can be characterized by

$$\begin{aligned}&\sqrt{n}(\hat{\theta}_{ML}-\theta_{ML,0})-\left[\mathrm{E}\left(-\frac{1}{n}\frac{\partial^2 \ln L_n\left(\theta_{ML,0}\right)}{\partial\theta_{ML}\partial\theta_{ML}'}\right)\right]^{-1}\Delta_n(\hat{\theta}_{ML})\\ &=\left[\mathrm{E}\left(-\frac{1}{n}\frac{\partial^2 \ln L_n\left(\theta_{ML,0}\right)}{\partial\theta_{ML}\partial\theta_{ML}'}\right)\right]^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_n^{\nu}(\theta_{ML,0})}{\partial\theta_{ML}}\xrightarrow{d} N\left(\mathbf{0},\mathcal{J}^{-1}(\theta_{ML,0})\right).\end{aligned} \tag{37}$$

The bias corrected 3SPMLE can be specified by

$$\hat{\theta}_{ML}^c=\hat{\theta}_{ML}-\frac{1}{n}\left[\mathrm{E}\left(-\frac{1}{n}\frac{\partial^2 \ln L_n\left(\theta_{ML,0}\right)}{\partial\theta_{ML}\partial\theta_{ML}'}\right)\right]^{-1}\Delta_n(\hat{\theta}_{ML}), \tag{38}$$

and we have $\sqrt{n}(\hat{\theta}_{ML}^c-\theta_{ML,0})\xrightarrow{d} N\left(\mathbf{0},\mathcal{J}^{-1}(\theta_{ML,0})\right)$.

6 Monte Carlo simulation

In this section, we conduct some Monte Carlo experiments to investigate the finite sample properties and the robustness of our proposed estimators. In the basic experiment design, x_i is a vector of dimension 3 with $x_{1,i}=1$. x_i^*, z_i^* and v_i^* follow a multivariate normal distribution

$$\begin{aligned}&\begin{pmatrix}x_i^*\\ z_i^*\\ v_i^*\end{pmatrix}\sim N\left(\begin{bmatrix}0\\0\\0\end{bmatrix},\ \begin{bmatrix}P_x & P_{xz} & \mathbf{0}_{k\times 1}\\ P_{xz}' & P_z & \rho_{vz}\\ \mathbf{0}_{k\times 1}' & \rho_{vz}' & 1\end{bmatrix}\right),\\ &x_{\tau,i}=R_\tau^{-1}(U_{x_{\tau,i}})=R_\tau^{-1}(\Phi(x_{\tau,i}^*)),\ z_{\iota,i}=H_\iota^{-1}(U_{z_{\iota,i}})=H_\iota^{-1}(\Phi(z_{\iota,i}^*)),\\ &v_i=G^{-1}(U_{v_i})=G^{-1}(\Phi(v_i^*))=\Phi^{-1}(\Phi(v_i^*))=v_i^*.\end{aligned} \tag{39}$$

For a SAR model with an endogenous weight matrix, $y_i=\lambda\sum_{j\neq i}w_{ij}y_j+x_i'\beta+v_i$ with a row-normalized weight matrix generated following Qu and Lee (2015), where $W_n=W_n^d\circ W_n^e$, i.e. $w_{ij}=w_{ij}^d w_{ij}^e$. $W_n^d=W_{R,n}^q$ is a predetermined queen-based contiguity weight matrix, where $W_{R,n}^q=(w_{R,ij}^q)$ with $w_{R,ij}^q=w_{ij}^q/(\sum_{k=1}^n w_{ik}^q)$. W_n^e is the endogenous part based on socio-economic similarity with $w_{ij}^e=1/|z_i-z_j|$ if $i\neq j$ and $w_{ii}^e=0$. For a SAR model with endogenous heterogeneity, we consider $y_i=\rho F(z_i'\varrho)\sum_{j\neq i}w_{ij}y_j+x_i'\beta+v_i$, where $F(\cdot)$ is the CDF for the Logistic distribution[27] with the location parameter $\mu=0$ and the scale parameter $s=1$. In this case, we have $W_n=W_n^d$. [28] Unless stated otherwise, the number of cross-sectional units $n=361$, the endogenous variables z_i only enter the model nonlinearly[29], and the true model settings assume moderate to high correlation between z_i and v_i, or z_i^* and v_i^*. The exact level of correlation varies with settings, as indicated by

27. As mentioned in Section 3.3.2, the IV approach can only be applied to a special case of this model, where the $F(\cdot)$ function doesn't include any parameters. Therefore, in the supplement file, we provide additional simulation results for a slightly different model $y_i=\varrho F(z_i)\sum_{j\neq i}w_{ij}y_j+x_i'\beta+v_i$, where the bounded function $F(z_i)$ is the same Logistic function.

28. The simulation results for the extension of endogenous heterogeneity with group fixed effect as in Section 5 are provided in the supplement file, the biased corrected estimator help to improve the finite sample estimation performance.

29. That is, z_i would only be part of the spatial weightings or the heterogeneity component but would not be regressors.

ρ_{vz} in the tables. 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)[30]. In the following tables, empirical standard errors are reported in parentheses, while square brackets contain the averaged asymptotic analytical errors ((A.1) or (A.2) in the appendix). Table 3 to Table 10 also report $p10$, $p30$, $p50$, $p70$ and $p90$, which represent the percentiles of differences between estimated parameters and the true values, illustrating the degree of symmetry and the range.

Different sample sizes. From Table 1 to Table 2, the performance of our estimators are evaluated through various sample sizes $n = 49,\ 144,\ 361,\ 529,\ 1024$. In this experiment, we consider a scalar z_i, which correlates with $x_{3,i}$ (and v_i). That is, both $x_i^*(= x_{3,i}^*)$ and z_i^* are scalars, and the underlying determination structure follows

$$\begin{pmatrix} x_i^* \\ z_i^* \\ v_i^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{bmatrix} \right). \tag{40}$$

As for regressors, we have $x_i = (x_{1,i}, x_{2,i}, x_{3,i})'$. $x_{1,i} = 1$, $x_{2,i} \sim N(0,1)$, and $x_{3,i} = R_3^{-1}(\Phi(x_{3,i}^*)) = \Phi_{(0,2^2)}^{-1}(\Phi(x_{3,i}^*))$. $z_i = H^{-1}(\Phi(z_i^*)) = \Phi^{-1}(\Phi(z_i^*)) = z_i^*$. Our estimators perform well across a series of sample sizes with correct specifications, and we detect that a sample size larger than or equal to 150 would be adequate for good finite sample performance. In Table 1, the IV estimator has a lower bias, especially in small to moderate sample sizes, but is less efficient than MLE.

Multiple x_i^* and z_i^*. Table 3 to Table 4 contain the case where we have multiple x_i^* and z_i^*. In this simulation, $z_i = (z_{1,i}, z_{2,i})'$, which correlates with $x_{2,i}$ and $x_{3,i}$ (and also v_i). $x_i^* = (x_{2,i}^*, x_{3,i}^*)'$, $z_i^* = (z_{1,i}^*, z_{2,i}^*)'$, and v_i^* are generated from

$$\begin{pmatrix} x_{2,i}^* \\ x_{3,i}^* \\ z_{1,i}^* \\ z_{2,i}^* \\ v_i^* \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0.5 & 0.5 & 0 \\ 0 & 1 & 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 1 & 0.5 & 0.25 \\ 0.5 & 0.5 & 0.5 & 1 & 0.25 \\ 0 & 0 & 0.25 & 0.25 & 1 \end{bmatrix} \right). \tag{41}$$

We have $x_i = (x_{1,i}, x_{2,i}, x_{3,i})'$, where $x_{1,i} = 1$, $x_{2,i} = R_2^{-1}(\Phi(x_{2,i}^*)) = \Phi^{-1}(\Phi(x_{2,i}^*)) = x_{2,i}^*$, $x_{3,i} = R_3^{-1}(\Phi(x_{3,i}^*)) = \Phi_{(0,2^2)}^{-1}(\Phi(x_{3,i}^*))$. $z_{\iota,i} = H^{-1}(\Phi(z_{\iota,i}^*)) = \Phi^{-1}(\Phi(z_{\iota,i}^*)) = z_{\iota,i}^*$, $\iota \in \{1,2\}$. We also compare our estimator with several other methods. In Table 3, SAR refers to the SAR model which incorrectly assumes no endogeneity in W_n. Original is the estimator proposed in the seminal work of Park and Gupta (2012), which, according to Yang et al. (2022) and as shown in Section S.1 in the supplement file, would be inconsistent due to the correlation between x_i^* and z_i^*. Thus, we expect both of them to display some bias. Heterogeneity in Table 4 indicates a model with continuous heterogeneity $\rho F(z_i'\varrho)$, but mistakenly treat z_i as exogenous. As expected, the misspecified SAR, Original, and Heterogeneity display considerable bias, while our MLE or IV estimator successfully corrects the endogeneity. In

30. 95% CP represents the proportion of the 95% asymptotic-distribution-based confidence intervals that contain the true parameters.

contrast, Table 5 contains the result for the case where there is no endogenous in the spatial weight matrix in the true data generating process, all the estimators for the SAR, Original, our MLE and IV are well-behaved.

z_i **also added as linear regressors.** Table 6 shows the result with endogenous z_i also serving as a linear regressor outside the spatial interaction term. We follow the setting as in Table 1, and have z_i as part of linear regressors. Yang et al. (2022) indicates that at least one of x_i needs to be non-Normal for identification since otherwise, there is multicollinearity. Here we still have a Normally distributed x_i^* that enters into the determination of z_i^*, but the correlated $x_{3,i} = R_3^{-1}(U_{x_{3,i}}) = R_3^{-1}(\Phi(x_{3,i}^*))$ needs to be non-Normal. Table 6 illustrates the results when we consider the CDF of Exponential (with mean 2) or Gamma(0.5,2) distribution [31] for R_3. We find that all other parameters are precisely estimated, except for a slight bias in β_z. A larger sample size is needed to guarantee a better performance.

Misspecifications. We first investigate the robustness of our estimators against the misspecifications of the joint distribution of v_i^* and z_i^*, which leads to the underlying endogeneity. Three non-Normal joint distributions for (v_i^*, z_i^*) are considered: (1) Wishart distribution [32], (2) t distribution [33], and (3) Gaussian Mixture distribution [34]. We set $z_i = \Phi^{-1}(\hat{H}(z_i^* + 0.5 \times x_{3,i}^*))$, where $\hat{H}$ is the empirical CDF calculated as in Section 3.1, $x_{3,i}^* \sim N(0,1)$, and $x_{3,i} = R_3^{-1}(\Phi(x_{3,i}^*)) = \Phi_{(0,2^2)}(\Phi(x_{3,i}^*))$. Table 7 and 8 illustrate that the most estimates are still robust [35], though the situation may deteriorate in the endogenous heterogeneity case with Wishart distribution, where the estimated ϱ may have a large bias.

Then we examine another type of misspecification. The baseline data is still generated by equation (40), but now $z_i = H^{-1}(\Phi(z_i^*)) + 0.5 \times \varepsilon_{1,i}^2 + 10 \times \varepsilon_{2,i} + v_i^*$, where H is the CDF for the exponential

31. Although theoretically, any non-Normal distributed $x_{3,i}$ would satisfy the condition, in practice, for a small to moderate sample size, the distribution of $x_{3,i}$ needs to be very distinct from Normal. For instance, $Gamma(0.5, 2)$ has a shape similar to exponential, which is completely different from Normal. However, the identification comes after the bell-shaped $Gamma(9, 0.5)$ could still be problematic. With high-level multicollinearity, the coefficients in front of z_i and its correlated $x_{3,i}$ would be influenced. So we recommend a large sample size to justify an instrumental-free estimation of linear parameters.

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

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

In the simulation, we set $d = 2$, $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, and $\nu = 4$. For each $i \in n$, we generate a random matrix and pick the diagonal elements as (v_i^*, z_i^*).

33. Multivariate t distribution for vector (v_i, ε_i) 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 $\nu = 3$.

34. The Gassian Mixture distribution takes the form

$$0.4 \times N\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right) + 0.4 \times N\left(\begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right) + 0.2 \times N\left(\begin{bmatrix} -8/3 \\ -8/3 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right).$$

35. ρ_{vz} are omitted here for simplicity, since there is no longer joint Normal distribution, so there is no true $\rho_{vz,0}$.

distribution with a rate parameter $\mu = 2$. $\varepsilon_{1,i}$ follows a Gaussian Mixture distribution [36] and $\varepsilon_{2,i} \sim U(0,1)$. In this case, we assume that we do not observe either $\varepsilon_{1,i}$, $\varepsilon_{2,i}$ or the exact functional form. If z_i enters only in a nonlinear manner, then even without excluded instruments, regressing z_i on x_i and inserting the error term as a control variable would not cause any identification problem [37]. However, this simple estimator is different from Copulas and is less robust when the functional form is not linear or the omitted variables are non-Normally distributed. Table 9 illustrates this. Both SAR-with IV and SAR-no IV refer to the estimator proposed in Qu and Lee (2015), where the potentially endogenous z_i is expressed and estimated as a linear function. The difference between them is whether or not we observe the excluded IV, $\varepsilon_{1,i}$ and $\varepsilon_{2,i}$, in the control function. We can see that despite the misspecification, our Copula estimator manages to grasp a reasonable estimate for structure parameters β_0, β_1, β_2 and λ. In contrast, Table 9 shows that the standard errors increase and the coverage deviates from 95% using Qu and Lee (2015)'s estimator, which indicates potential inference problems. Similar results apply to the heterogeneity case in Table 10, where EHSAR denotes the estimator proposed in Lin et al. (2022). Our MLE offers precise estimates in β and ρ and is much less biased in ϱ. However, this simulation just offers an example of the superior performance of our Copula estimator. Whether or not this can be supported theoretically leaves for further study.

36. The Gaussian Mixture is $0.4 \times N(1,1) + 0.4 \times N(1/3,1) + 0.2 \times N(-8/3,1)$.
37. Qu and Lee (2015) also mentioned this.

			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	0.9913	0.9671	0.9974	0.9925	0.9979	0.9950	0.9976	0.9954	0.9968	0.9960
		Std	(0.1996)	(0.2144)	(0.0995)	(0.1039)	(0.0597)	(0.0624)	(0.0472)	(0.0494)	(0.0354)	(0.0378)
			[0.1385]	[0.1690]	[0.0697]	[0.0892]	[0.0475]	[0.0544]	[0.0423]	[0.0446]	[0.0310]	[0.0320]
		Coverage	0.9510	0.9470	0.9480	0.9530	0.9420	0.9420	0.9530	0.9510	0.9470	0.9490
β_1	4	Mean	4.0035	3.9986	4.0007	3.9995	4.0022	4.0012	4.0000	3.9994	3.9994	3.9991
		Std	(0.1238)	(0.1238)	(0.0728)	(0.0740)	(0.0451)	(0.0457)	(0.0364)	(0.0369)	(0.0255)	(0.0256)
			[0.0977]	[0.1223]	[0.0689]	[0.0707]	[0.0447]	[0.0445]	[0.0374]	[0.0368]	[0.0259]	[0.0264]
		Coverage	0.9460	0.9460	0.9550	0.9570	0.9500	0.9460	0.9480	0.9430	0.9420	0.9500
β_2	-2	Mean	-2.0018	-1.9980	-2.0028	-2.0019	-2.0010	-2.0003	-2.0005	-2.0001	-1.9997	-1.9995
		Std	(0.0785)	(0.0795)	(0.0439)	(0.0441)	(0.0270)	(0.0276)	(0.0223)	(0.0226)	(0.0160)	(0.0162)
			[0.0523]	[0.0750]	[0.0326]	[0.0433]	[0.0215]	[0.0274]	[0.0181]	[0.0226]	[0.0130]	[0.0162]
		Coverage	0.9490	0.9480	0.9500	0.9590	0.9520	0.9530	0.9540	0.9500	0.9450	0.9470
λ	0.8	Mean	0.7974	0.8013	0.7990	0.7999	0.7997	0.8003	0.7999	0.8003	0.8001	0.8002
		Std	(0.0182)	(0.0210)	(0.0091)	(0.0107)	(0.0053)	(0.0065)	(0.0044)	(0.0053)	(0.0032)	(0.0040)
			[0.0170]	[0.0191]	[0.0081]	[0.0104]	[0.0054]	[0.0064]	[0.0046]	[0.0053]	[0.0033]	[0.0038]
		Coverage	0.9470	0.9420	0.9440	0.9420	0.9490	0.9450	0.9560	0.9470	0.9440	0.9550
ρ_{vz}	0.5	Mean	0.4802	-	0.4925	-	0.4962	-	0.4964	-	0.4987	-
		Std	(0.0977)	-	(0.0552)	-	(0.0357)	-	(0.0298)	-	(0.0211)	-
			[0.1100]	-	[0.0654]	-	[0.0415]	-	[0.0343]	-	[0.0247]	-
		Coverage	0.9410	-	0.9450	-	0.9450	-	0.9480	-	0.9490	-
σ_v	1	Mean	0.9635	-	0.9878	-	0.9948	-	0.9958	-	0.9978	-
		Std	(0.1025)	-	(0.0596)	-	(0.0371)	-	(0.0298)	-	(0.0215)	-
			[0.0977]	-	[0.0583]	-	[0.0371]	-	[0.0307]	-	[0.0221]	-
		Coverage	0.9320	-	0.9500	-	0.9480	-	0.9450	-	0.9510	-

Table 1: Estimates from endogenous spatial weight matrices under different sample sizes.

			n = 49	n = 144	n = 361	n = 529	n = 1024
			MLE	MLE	MLE	MLE	MLE
β_0	1	Mean	0.9850	0.9971	0.9971	0.9971	0.9969
		Std	(0.1931)	(0.0990)	(0.0592)	(0.0470)	(0.0356)
			[0.1351]	[0.0682]	[0.0477]	[0.0426]	[0.0314]
		Coverage	0.9600	0.9520	0.9470	0.9450	0.9430
β_1	4	Mean	3.9989	3.9997	4.0018	3.9996	3.9995
		Std	(0.1250)	(0.0720)	(0.0445)	(0.0356)	(0.0250)
			[0.1007]	[0.0676]	[0.0441]	[0.0369]	[0.0256]
		Coverage	0.9430	0.9560	0.9460	0.9460	0.9460
β_2	-2	Mean	-2.0032	-2.0032	-2.0008	-2.0006	-1.9998
		Std	(0.0801)	(0.0445)	(0.0271)	(0.0221)	(0.0158)
			[0.0516]	[0.0320]	[0.0216]	[0.0183]	[0.0132]
		Coverage	0.9510	0.9540	0.9540	0.9510	0.9430
ρ	0.8	Mean	0.7878	0.7945	0.7990	0.7996	0.8002
		Std	(0.1001)	(0.0518)	(0.0330)	(0.0271)	(0.0193)
			[0.1056]	[0.0454]	[0.0345]	[0.0286]	[0.0210]
		Coverage	0.9420	0.9500	0.9530	0.9500	0.9550
ϱ	0.5	Mean	0.6345	0.5263	0.5054	0.5061	0.5019
		Std	(1.5006)	(0.1692)	(0.0964)	(0.0821)	(0.0573)
			[0.4706]	[0.1573]	[0.1057]	[0.0811]	[0.0551]
		Coverage	0.9980	0.9500	0.9520	0.9490	0.9510
ρ_{vz}	0.5	Mean	0.4751	0.4915	0.4963	0.4962	0.4986
		Std	(0.1142)	(0.0601)	(0.0376)	(0.0325)	(0.0226)
			[0.1100]	[0.0654]	[0.0415]	[0.0343]	[0.0247]
		Coverage	0.9410	0.9440	0.9500	0.9550	0.9440
σ_v	1	Mean	0.9537	0.9843	0.9938	0.9950	0.9974
		Std	(0.1136)	(0.0642)	(0.0393)	(0.0321)	(0.0229)
			[0.0965]	[0.0580]	[0.0370]	[0.0306]	[0.0221]
		Coverage	0.9320	0.9470	0.9540	0.9550	0.9530

Table 2: Estimates from endogenous heterogeneity under different sample sizes.

		True	Mean	Std	Coverage	p10	p30	p50	p70	p90
β_0	SAR	1	0.4838	(1.0741)	0.8990	-2.1239	-0.0947	-0.0304	0.0187	0.0787
	Original		0.9925	(0.0761)	0.9460	-0.1096	-0.0467	-0.0086	0.0299	0.0921
	MLE		1.0006	(0.0666) [0.6000]	0.9560	-0.0834	-0.0338	0.0017	0.0371	0.0859
	IV		0.9959	(0.0708) [0.0651]	0.9470	-0.0939	-0.0403	-0.0044	0.0342	0.0856
β_1	SAR	4	3.8605	(0.3375)	0.9390	-0.5528	-0.0754	-0.0196	0.0141	0.0557
	Original		3.5047	(0.0619)	0.0000	-0.5783	-0.5290	-0.4918	-0.4615	-0.4181
	MLE		3.9971	(0.0529) [0.0463]	0.9380	-0.0714	-0.0280	-0.0021	0.0239	0.0652
	IV		3.9960	(0.0533) [0.0537]	0.9410	-0.0730	-0.0305	-0.0031	0.0223	0.0640
β_2	SAR	-2	-1.9399	(0.1496)	0.9400	-0.0296	-0.0084	0.0104	0.0332	0.2356
	Original		-2.2491	(0.0315)	0.0000	-0.2899	-0.2658	-0.2487	-0.2337	-0.2099
	MLE		-2.0017	(0.0258) [0.0231]	0.9500	-0.0342	-0.0159	-0.0011	0.0123	0.0311
	IV		-2.0012	(0.0259) [0.0267]	0.9550	-0.0335	-0.0151	-0.0010	0.0131	0.0319
λ	SAR	0.8	0.9042	(0.2121)	0.9040	-0.0104	-0.0022	0.0034	0.0119	0.4060
	Original		0.7995	(0.0078)	0.9490	-0.0096	-0.0045	-0.0006	0.0036	0.0092
	MLE		0.7995	(0.0078) [0.0075]	0.9510	-0.0097	-0.0045	-0.0007	0.0035	0.0093
	IV		0.8004	(0.0090) [0.0090]	0.9450	-0.0109	-0.0041	0.0003	0.0047	0.0117
ρ_{vz_1}	SAR	0.25	-	-	-	-	-	-	-	-
	Original		0.6102	(0.0398)	0.0000	0.3084	0.3420	0.3602	0.3819	0.4085
	MLE		0.2492	(0.0352) [0.0356]	0.9530	-0.0454	-0.0200	0.0006	0.0176	0.0450
	IV		-	-	-	-	-	-	-	-
ρ_{vz_2}	SAR	0.25	-	-	-	-	-	-	-	-
	Original		0.6095	(0.0392)	0.0000	0.3099	0.3416	0.3608	0.3802	0.4088
	MLE		0.2473	(0.0339) [0.0355]	0.9530	-0.0488	-0.0194	-0.0023	0.0148	0.0408
	IV		-	-	-	-	-	-	-	-
σ_v	SAR	1	2.9105	(5.2318)	0.9440	-0.0976	-0.0378	0.0134	0.0922	6.1358
	Original		1.2172	(0.0620)	0.0560	0.1432	0.1822	0.2155	0.2471	0.3027
	MLE		0.9926	(0.0370) [0.0370]	0.9500	-0.0536	-0.0277	-0.0083	0.0123	0.0408
	IV		-	-	-	-	-	-	-	-

Table 3: Estimates from endogenous spatial weight matrices with multiple x_i^* and z_i^*. n = 361.

		True	Mean	Std	Coverage	p10	p30	p50	p70	p90
β_0	SAR	1	1.0039	(0.0698)	0.9500	-0.0844	-0.0328	0.0016	0.0377	0.0984
	Heterogeneity		0.9994	(0.0620)	0.9500	-0.0815	-0.0337	-0.0020	0.0307	0.0780
	Original		0.9932	(0.0677)	0.9500	-0.0943	-0.0433	-0.0083	0.0266	0.0799
	MLE		1.0026	(0.0591) [0.0538]	0.9600	-0.0729	-0.0301	0.0008	0.0347	0.0790
β_1	SAR	4	3.9985	(0.0620)	0.9550	-0.0834	-0.0345	-0.0008	0.0335	0.0761
	Heterogeneity		3.9439	(0.0657)	0.8660	-0.1438	-0.0917	-0.0543	-0.0222	0.0265
	Original		3.5047	(0.0676)	0.0000	-0.5810	-0.5299	-0.4964	-0.4588	-0.4084
	MLE		3.9960	(0.0578) [0.0480]	0.9470	-0.0786	-0.0327	-0.0050	0.0256	0.0671
β_2	SAR	-2	-2.0003	(0.0313)	0.9530	-0.0400	-0.0168	-0.0016	0.0170	0.0423
	Heterogeneity		-2.0279	(0.0321)	0.8540	-0.0696	-0.0456	-0.0274	-0.0096	0.0135
	Original		-2.2484	(0.0329)	0.0000	-0.2900	-0.2664	-0.2492	-0.2312	-0.2064
	MLE		-2.0024	(0.0284) [0.0241]	0.9540	-0.0394	-0.0180	-0.0012	0.0136	0.0336
ρ	SAR	0.8	0.4005	(0.0254)	0.0000	-0.4321	-0.4128	-0.3997	-0.3849	-0.3661
	Heterogeneity		0.8008	(0.0380)	0.9540	-0.0494	-0.0190	0.0018	0.0214	0.0485
	Original		0.7981	(0.0332)	0.9580	-0.0441	-0.0209	-0.0018	0.0174	0.0394
	MLE		0.7982	(0.0331) [0.0334]	0.9570	-0.0442	-0.0211	-0.0019	0.0179	0.0391
ϱ_1	Heterogeneity	1	1.2144	(0.1960)	0.0000	-0.0306	0.1052	0.2057	0.3106	0.4768
	Original		1.0186	(0.1662)	0.9530	-0.1825	-0.0728	0.0077	0.0924	0.2312
	MLE		1.0200	(0.1664) [0.1469]	0.9510	-0.1846	-0.0717	0.0105	0.0930	0.2363
ϱ_2	Heterogeneity	-1	-0.8251	(0.1634)	0.0000	-0.0372	0.1026	0.1875	0.2680	0.3653
	Original		-1.0119	(0.1653)	0.9480	-0.2282	-0.0882	0.0002	0.0775	0.1894
	MLE		-1.0096	(0.1654) [0.1461]	0.9510	-0.2232	-0.0836	0.0011	0.0826	0.1941
ρ_{vz_1}	Original	0.25	0.4064	(0.0485)	0.0990	0.0953	0.1309	0.1567	0.1814	0.2207
	MLE		0.2466	(0.0370) [0.0353]	0.9500	-0.0497	-0.0235	-0.0034	0.0154	0.0445
ρ_{vz_2}	Original	0.25	0.4080	(0.0510)	0.1270	0.0923	0.1333	0.1620	0.1871	0.2191
	MLE		0.2476	(0.0395) [0.0353]	0.9430	-0.0558	-0.0220	-0.0020	0.0196	0.0480
σ_v	SAR	1	1.2772	(0.1178)	0.3670	0.1349	0.2161	0.2647	0.3265	0.4391
	Heterogeneity		0.9524	(0.0705)	0.8920	-0.1383	-0.0866	-0.0507	-0.0097	0.0433
	Original		1.2157	(0.0674)	0.1160	0.1285	0.1784	0.2147	0.2507	0.3030
	MLE		0.9902	(0.0367) [0.0369]	0.9470	-0.0569	-0.0299	-0.0108	0.0094	0.0390

Table 4: Estimates from endogenous heterogeneity with multiple x_i^* and z_i^*. n = 361.

		True	Mean	Std	Coverage	p10	p30	p50	p70	p90
β_0	SAR	1	1.0023	(0.0643)	0.9480	-0.0754	-0.0324	0.0017	0.0334	0.0836
	Original		1.0025	(0.0649)	0.9500	-0.0776	-0.0325	0.0020	0.0349	0.0840
	MLE		1.0025	(0.0648) [0.0581]	0.9500	-0.0773	-0.0325	0.0021	0.0349	0.0845
	IV		0.9982	(0.0686) [0.0667]	0.9480	-0.0863	-0.0384	-0.0030	0.0310	0.0881
β_1	SAR	4	4.0027	(0.0551)	0.9500	-0.0718	-0.0231	0.0051	0.0324	0.0715
	Original		4.0027	(0.0552)	0.9510	-0.0721	-0.0227	0.0053	0.0326	0.0713
	MLE		4.0027	(0.0552) [0.0547]	0.9510	-0.0720	-0.0223	0.0052	0.0324	0.0714
	IV		4.0011	(0.0560) [0.0545]	0.9450	-0.0748	-0.0256	0.0035	0.0319	0.0694
β_2	SAR	-2	-2.0009	(0.0274)	0.9540	-0.0368	-0.0153	-0.0010	0.0138	0.0343
	Original		-2.0009	(0.0308)	0.9460	-0.0090	-0.0035	-0.0002	0.0030	0.0077
	MLE		-2.0009	(0.0275) [0.0263]	0.9540	-0.0372	-0.0153	-0.0008	0.0139	0.0341
	IV		-1.9998	(0.0283) [0.0282]	0.9530	-0.0369	-0.0150	0.0002	0.0146	0.0366
λ	SAR	0.8	0.7996	(0.0064)	0.9470	-0.0087	-0.0034	-0.0002	0.0029	0.0073
	Original		0.7996	(0.0065)	0.9450	-0.0090	-0.0035	-0.0002	0.0030	0.0077
	MLE		0.7996	(0.0065) [0.0066]	0.9460	-0.0573	-0.0244	-0.0006	0.0245	0.0590
	IV		0.8005	(0.0079) [0.0079]	0.9470	-0.0092	-0.0034	0.0001	0.0045	0.0106
ρ_{vz}	SAR	0	-	-	-	-	-	-	-	-
	Original		0.0004	(0.0611)	0.9530	-0.0774	-0.0323	-0.0008	0.0326	0.0785
	MLE		0.0004	(0.0453) [0.0454]	0.9580	-0.0573	-0.0244	-0.0006	0.0245	0.0590
	IV		-	-	-	-	-	-	-	-
σ_v	SAR	1	0.9908	(0.0728)	0.9470	-0.1014	-0.0499	-0.0091	0.0284	0.0848
	Original		0.9953	(0.0366)	0.9490	-0.0516	-0.0248	-0.0040	0.0144	0.0421
	MLE		0.9948	(0.0366) [0.0371]	0.9480	-0.0522	-0.0252	-0.0046	0.0143	0.0417
	IV		-	-	-	-	-	-	-	-

Table 5: Estimates when there's no endogeneity in the spatial weight matrix. n = 361.

			True	Mean	Std	Coverage	p10	p30	p50	p70	p90
β_0	Gamma	MLE	1	1.0057	(0.0829) [0.0639]	0.9550	-0.1026	-0.0372	0.0044	0.0540	0.1110
		IV		1.0092	(0.0849) [0.0769]	0.9500	-0.1019	-0.0340	0.0081	0.0571	0.1173
	Exponential	MLE		1.0218	(0.1358) [0.1001]	0.9510	-0.1515	-0.0464	0.0264	0.0968	0.1921
		IV		1.0309	(0.1438) [0.1357]	0.9360	-0.1543	-0.0352	0.0326	0.1051	0.2097
β_1	Gamma	MLE	4	4.0020	(0.0453) [0.0447]	0.9510	-0.0583	-0.0193	0.0032	0.0259	0.0590
		IV		4.0004	(0.0460) [0.0448]	0.9460	-0.0603	-0.0222	0.0025	0.0257	0.0584
	Exponential	MLE		4.0019	(0.0457) [0.0445]	0.9470	-0.0582	-0.0183	0.0033	0.0270	0.0589
		IV		4.0007	(0.0457) [0.0444]	0.9440	-0.0600	-0.0203	0.0030	0.0258	0.0577
β_2	Gamma	MLE	-2	-2.0103	(0.0581) [0.0317]	0.9450	-0.0840	-0.0405	-0.0111	0.0164	0.0651
		IV		-2.0095	(0.0582) [0.0564]	0.9440	-0.0836	-0.0393	-0.0116	0.0175	0.0657
	Exponential	MLE		-2.0135	(0.0517) [0.0229]	0.9430	-0.0785	-0.0404	-0.0150	0.0114	0.0528
		IV		-2.0129	(0.0517) [0.0504]	0.9420	-0.0781	-0.0404	-0.0143	0.0118	0.0537
β_z	Gamma	MLE	-1	-0.9702	(0.1686) [0.0454]	0.9500	-0.1891	-0.0574	0.0346	0.1237	0.2325
		IV		-0.9687	(0.1694) [0.1665]	0.9480	-0.1868	-0.0601	0.0362	0.1263	0.2336
	Exponential	MLE		-0.9454	(0.2095) [0.0460]	0.9460	-0.2080	-0.0508	0.0661	0.1724	0.3172
		IV		-0.9440	(0.2102) [0.2060]	0.9490	-0.2119	-0.0507	0.0670	0.1746	0.3135
λ	Gamma	MLE	0.8	0.7998	(0.0060) [0.0059]	0.9470	-0.0078	-0.0030	-0.0001	0.0028	0.0072
		IV		0.8007	(0.0072) [0.0071]	0.9410	-0.0079	-0.0029	0.0005	0.0045	0.0099
	Exponential	MLE		0.7999	(0.0052) [0.0051]	0.9510	-0.0070	-0.0026	-0.0001	0.0024	0.0064
		IV		0.8006	(0.0062) [0.0061]	0.9420	-0.0068	-0.0026	0.0006	0.0037	0.0082
ρ_{vz}	Gamma	MLE	0.5	0.4715	(0.0979) [0.0417]	0.9450	-0.1529	-0.0749	-0.0220	0.0293	0.0887
		IV		-	-	-	-	-	-	-	-
	Exponential	MLE		0.4525	(0.1216) [0.0419]	0.9360	-0.2089	-0.1008	-0.0394	0.0252	0.0953
		IV		-	-	-	-	-	-	-	-
σ_v	Gamma	MLE	1	0.9863	(0.0884) [0.0368]	0.9550	-0.1185	-0.0641	-0.0246	0.0246	0.1011
		IV		-	-	-	-	-	-	-	-
	Exponential	MLE		0.9788	(0.1037) [0.0365]	0.9590	-0.1404	-0.0818	-0.0374	0.0213	0.1139
		IV		-	-	-	-	-	-	-	-

Table 6: Estimates from endogenous spatial weight matrix when z_i also enters as a regressor. n = 361.

			True	Mean	Std	Coverage	p10	p30	p50	p70	p90
β_0	Wishart	MLE	1	1.0096	(0.0726) [0.1543]	0.9430	-0.0747	-0.0227	0.0057	0.0373	0.0947
		IV		0.9906	(0.1018) [0.1735]	0.9410	-0.1245	-0.0548	-0.0125	0.0322	0.1102
	t	MLE		1.0095	(0.1099) [0.0973]	0.9580	-0.1249	-0.0481	0.0031	0.0595	0.1483
		IV		0.9801	(0.1136) [0.1016]	0.9410	-0.1526	-0.0754	-0.0249	0.0351	0.1256
	Mixture	MLE		0.9754	(0.1180) [0.0859]	0.9510	-0.1734	-0.0884	-0.0291	0.0329	0.1306
		IV		0.9753	(0.1213) [0.0985]	0.9440	-0.1765	-0.0870	-0.0282	0.0360	0.1303
β_1	Wishart	MLE	4	4.0042	(0.1503) [0.1545]	0.9510	-0.1923	-0.0710	0.0104	0.0888	0.1917
		IV		3.9948	(0.1538) [0.1491]	0.9530	-0.2126	-0.0805	0.0016	0.0755	0.0755
	t	MLE		4.0044	(0.0834) [0.0830]	0.9500	-0.0976	-0.0365	0.0008	0.0443	0.1141
		IV		3.9947	(0.0842) [0.0826]	0.9490	-0.1098	-0.0467	-0.0090	0.0371	0.1028
	Mixture	MLE		3.9954	(0.0718) [0.0733]	0.9480	-0.0975	-0.0404	-0.0024	0.0341	0.0853
		IV		3.9958	(0.0724) [0.0736]	0.9490	-0.0970	-0.0414	-0.0022	0.0355	0.0856
β_2	Wishart	MLE	-2	-2.0024	(0.0749) [0.0714]	0.9570	-0.0985	-0.0434	0.0001	0.0364	0.0934
		IV		-1.9974	(0.0755) [0.0767]	0.9510	-0.0963	-0.0374	0.0070	0.0433	0.0981
	t	MLE		-2.0067	(0.0461) [0.0438]	0.9450	-0.0687	-0.0268	-0.0035	0.0161	0.0491
		IV		-2.0007	(0.0459) [0.0465]	0.9450	-0.0620	-0.0214	0.0009	0.0224	0.0577
	Mixture	MLE		-1.9976	(0.0474) [0.0363]	0.9500	-0.0573	-0.0224	0.0004	0.0274	0.0654
		IV		-1.9977	(0.0477) [0.0491]	0.9510	-0.0602	-0.0238	0.0007	0.0279	0.0640
λ	Wishart	MLE	0.8	0.7963	(0.0166) [0.0152]	0.9450	-0.0250	-0.0116	-0.0031	0.0058	0.0170
		IV		0.8019	(0.0232) [0.0225]	0.9490	-0.0283	-0.0095	0.0037	0.0148	0.0292
	t	MLE		0.7968	(0.0112) [0.0098]	0.9490	-0.0158	-0.0075	-0.0024	0.0022	0.0088
		IV		0.8026	(0.0125) [0.0119]	0.9460	-0.0128	-0.0027	0.0031	0.0083	0.0166
	Mixture	MLE		0.8026	(0.0089) [0.0093]	0.9420	-0.0091	-0.0019	0.0027	0.0071	0.0143
		IV		0.8024	(0.0113) [0.0110]	0.9550	-0.0124	-0.0037	0.0024	0.0083	0.0168
σ_v	Wishart	MLE	2.8227	2.8122	(0.1661) [0.1063]	0.9490	-0.2183	-0.1043	-0.0138	0.0638	0.2066
	t	MLE	1.7467	1.6964	(0.3460) [0.0636]	0.9730	-0.3113	-0.1981	-0.1147	-0.0039	0.2071
	Mixture	MLE	1.8417	1.8220	(0.0705) [0.0682]	0.9350	-0.1135	-0.0584	-0.0203	0.0190	0.0691

Table 7: Estimates from non-Normal joint distributions of z_i^* and v_i^* with an endogenous weight matrix. n = 361.

		True	Mean	Std	Coverage	p10	p30	p50	p70	p90
β_0	Wishart	1	1.0066	(0.0805) [0.1517]	0.9470	-0.0914	-0.0352	0.0031	0.0437	0.1116
	t		0.9944	(0.1016) [0.0950]	0.9540	-0.1314	-0.0630	-0.0074	0.0463	0.1267
	Mixture		0.9977	(0.1110) [0.0864]	0.9460	-0.1436	-0.0635	-0.0048	0.0526	0.1467
β_1	Wishart	4	4.0001	(0.1485) [0.1526]	0.9530	-0.1958	-0.0768	0.0111	0.0802	0.1817
	t		3.9985	(0.0821) [0.0803]	0.9530	-0.1011	-0.0400	-0.0047	0.0384	0.1059
	Mixture		4.0005	(0.0718) [0.0728]	0.9500	-0.0899	-0.0382	0.0022	0.0394	0.0917
β_2	Wishart	-2	-2.0014	(0.0741) [0.0705]	0.9560	-0.0952	-0.0407	0.0015	0.0368	0.0935
	t		-2.0018	(0.0468) [0.0433]	0.9450	-0.0645	-0.0233	0.0005	0.0212	0.0541
	Mixture		-1.9990	(0.0475) [0.0367]	0.9440	-0.0596	-0.0238	-0.0010	0.0257	0.0633
ρ	Wishart	0.8	0.7860	(0.0938) [0.0834]	0.9440	-0.1344	-0.0622	-0.0093	0.0370	0.1030
	t		0.7983	(0.0574) [0.0571]	0.9420	-0.0737	-0.0300	0.0006	0.0303	0.0669
	Mixture		0.7943	(0.0509) [0.0565]	0.9490	-0.0721	-0.0333	-0.0069	0.0230	0.0595
ϱ	Wishart	0.5	0.6108	(0.3704) [0.3073]	0.9440	-0.2919	-0.0851	0.0751	0.2317	0.5421
	t		0.4719	(0.2353) [0.1511]	0.9560	-0.3104	-0.1376	-0.0058	0.0969	0.2501
	Mixture		0.4926	(0.1767) [0.1646]	0.9530	-0.2224	-0.1112	-0.0215	0.0680	0.2237
σ_v	Wishart	2.8227	2.8016	(0.1654) [0.1046]	0.9420	-0.2273	-0.1129	-0.0256	0.0513	0.2017
	t	1.7467	1.6966	(0.3446) [0.0632]	0.9730	-0.3096	-0.1997	-0.1143	-0.0083	0.2106
	Mixture	1.8417	1.8283	(0.0758) [0.0681]	0.9400	-0.1120	-0.0544	-0.0130	0.0284	0.0820

Table 8: Estimates from non-Normal joint distributions of z_i^* and v_i^* with endogenous heterogeneity. n = 361.

		True	Mean	Std	Coverage	p10	p30	p50	p70	p90
β_0	SAR - with IV	1	1.0016	(0.0599)	0.9460	-0.0699	-0.0283	0.0006	0.0324	0.0768
	SAR – no IV		1.0010	(0.0883)	0.9900	-0.0760	-0.0305	0.0003	0.0338	0.0788
	MLE		0.9981	(0.0621) [0.0520]	0.9470	-0.0776	-0.0333	-0.0030	0.0300	0.0745
	IV		0.9939	(0.0661) [0.0611]	0.9420	-0.0878	-0.0398	-0.0053	0.0251	0.0789
β_1	SAR - with IV	4	4.0028	(0.0558)	0.9460	-0.0672	-0.0251	0.0023	0.0315	0.0741
	SAR – no IV		4.0026	(0.0691)	0.9810	-0.0681	-0.0255	0.0022	0.0321	0.0741
	MLE		4.0029	(0.0490) [0.0489]	0.9440	-0.0570	-0.0233	0.0026	0.0289	0.0645
	IV		4.0016	(0.0494) [0.0487]	0.9430	-0.0599	-0.0235	-0.0006	0.0295	0.0642
β_2	SAR - with IV	-2	-2.0005	(0.0272)	0.9500	-0.0350	-0.0142	-0.0001	0.0126	0.0346
	SAR – no IV		-2.0005	(0.0293)	0.9670	-0.0354	-0.0144	-0.0005	0.0122	0.0359
	MLE		-2.0003	(0.0268) [0.0235]	0.9480	-0.0351	-0.0145	-0.0005	0.0123	0.0349
	IV		-1.9997	(0.0271) [0.0271]	0.9520	-0.0347	-0.0143	-0.0007	0.0136	0.0377
λ	SAR - with IV	0.8	0.7994	(0.0052)	0.9450	-0.0073	-0.0033	-0.0004	0.0020	0.0060
	SAR – no IV		0.7997	(0.0090)	0.9900	-0.0080	-0.0033	-0.0003	0.0025	0.0069
	MLE		0.7996	(0.0059) [0.0059]	0.9450	-0.0079	-0.0033	-0.0003	0.0026	0.0069
	IV		0.8003	(0.0073) [0.0071]	0.9530	-0.0089	-0.0034	0.0005	0.0040	0.0096
σ_v	SAR - with IV	1	1.0021	(0.0625)	0.9679	-0.0542	-0.0227	-0.0021	0.0177	0.0542
	SAR – no IV		0.9939	(0.0438)	0.9770	-0.0540	-0.0242	-0.0052	0.0141	0.0422
	MLE		0.9946	(0.0369) [0.0371]	0.9470	-0.0532	-0.0240	-0.0051	0.0141	0.0417
	IV		-	-	-	-	-	-	-	-

Table 9: Estimates from linear model misspecification with an endogenous spatial weight matrix. n = 361.

		True	Mean	Std	Coverage	p10	p30	p50	p70	p90
β_0	EHSAR - with IV	1	0.9788	(0.1594)	0.9890	-0.0922	-0.0416	-0.0059	0.0278	0.0738
	EHSAR – no IV		0.9956	(0.0704)	0.9500	-0.0944	-0.0385	-0.0051	0.0314	0.0781
	MLE		0.9943	(0.0702) [0.0578]	0.9540	-0.0946	-0.0398	-0.0060	0.0301	0.0763
β_1	EHSAR - with IV	4	3.9978	0.0632	0.9640	-0.0718	-0.0281	-0.0007	0.0290	0.0700
	EHSAR – no IV		4.0011	(0.0557)	0.9440	-0.0687	-0.0266	0.0015	0.0311	0.0715
	MLE		4.0019	(0.0487) [0.0485]	0.9500	-0.0593	-0.0236	0.0013	0.0291	0.0617
β_2	EHSAR - with IV	-2	-1.9980	0.0333	0.9760	-0.0344	-0.0136	0.0001	0.0131	0.0370
	EHSAR – no IV		-1.9975	(0.0268)	0.9480	-0.0322	-0.0115	0.0028	0.0154	0.0373
	MLE		-1.9980	(0.0268) [0.0233]	0.9500	-0.0330	-0.0117	0.0024	0.0144	0.0372
ρ	EHSAR - with IV	0.8	0.8040	(0.0427)	0.9870	-0.0166	-0.0064	-0.0003	0.0061	0.0160
	EHSAR – no IV		0.7915	(0.0139)	0.9020	-0.0267	-0.0158	-0.0084	-0.0014	0.0086
	MLE		0.7942	(0.0148) [0.0162]	0.9240	-0.0248	-0.0133	-0.0058	0.0013	0.0128
ϱ	EHSAR - with IV	0.5	0.2943	(0.0390)	0.0050	-0.2435	-0.2224	-0.2079	-0.1936	-0.1676
	EHSAR – no IV		0.3321	(0.0455)	0.0510	-0.2212	-0.1928	-0.1715	-0.1481	-0.1106
	MLE		0.5626	(0.0865) [0.0981]	0.9000	-0.0401	0.0147	0.0566	0.0985	0.1678
σ_v	EHSAR - with IV	1	1.0017	(0.1075)	0.9890	-0.0582	-0.0269	-0.0070	0.0119	0.0430
	EHSAR – no IV		1.0025	(0.0395)	0.9540	-0.0499	-0.0181	0.0043	0.0234	0.0534
	MLE		1.0026	(0.0398) [0.0374]	0.9550	-0.0497	-0.0177	0.0041	0.0230	0.0535

Table 10: Estimates from linear model misspecification with endogenous heterogeneity. n = 361.

7 Empirical application

In this section, we provide an example for empirical application on the studies of potential spatial spillovers in regional productivity. Works on this topic include a new regional development model in Gennaioli et al. (2013) and its spatial extension by Sanso-Navarro et al. (2017). Here we deviate from the contiguity-based exogenous spatial weights in Sanso-Navarro et al. (2017) and provide an illustration with a potentially endogenous spatial weight matrix. We combine the data set collected in and made-public by Gennaioli et al. (2013) with the location data from Sanso-Navarro et al. (2017) [38]. For more details regarding the data, readers can refer to these two papers. We use all the regional characteristics without a large fraction of missing values in Gennaioli et al. (2013) in this application [39], including (1) average temperature, (2) inverse distance to coast, (3) ln oil production per capita, (4) ln population, (5) years of education, (6) ln number of ethnic groups, (7) the probability of a country's residents speaking the same language, (8) the electricity power grid density, and (9) ln travel times.

The model we estimated is Eq.(2) with an endogenous W_n. The outcome Y_n is the regional productivity measured by ln income per capita, and the spatial weight matrix $W_n = W_n^d \circ W_n^e$ as in Section 6 with W_n^d being constructed by the Rook contiguity in Hoshino (2022) and $W_n^e = 1/|z_i - z_j|$ if $i \neq j$ and $w_{ii}^e = 0$, where z_i is the years of education. We believe that, beyond the physical distance, spillovers may also depend on socioeconomic statuses, like the similarities in education level, which is consistent with findings in Stoyanov and Zubanov (2012) and He et al. (2018). For instance, Stoyanov and Zubanov (2012) show that productivity spillovers occur when workers move out from a high-productive firm. We may expect workers to move to a region with similar socioeconomic conditions, for which education level serves as a proxy. Moreover, firms with high productivity may prefer a new site with similar characteristics when relocating. Furthermore, z_i is potentially endogenous, especially when the country fixed effects are not incorporated in [40].

To illustrate the performance of our proposed copula endogeneity correction method, we consider two settings. For Setting 1, we have variables (1) to (6) as regressors (as in Hoshino (2022)), with variables (7) - (9) serving as excluded instruments. Thus, the control function approach in Qu and Lee (2015) applies here. Moreover, we construct a model where all the variables enter as regressors as Setting 2. In this case, it would be hard to find excluded instruments. Consequently, Qu and Lee (2015)'s method may break down, but technically the Copula-based method would still work. For the copula estimation, we have the correlated regressors x_i^* generated from all the exogenous characteristics. As documented in Yang et al. (2022), if z_i also serves as a linear regressor, for the identification purpose, we need some of the variables, which enter into x_i^*, to be non-Normally distributed. Figure 1 demonstrates that this requirement for identification is satisfied since most variables follow distributions that are different from Normal.

Table 11 and 12 report the estimated results from different models, where all the estimators are

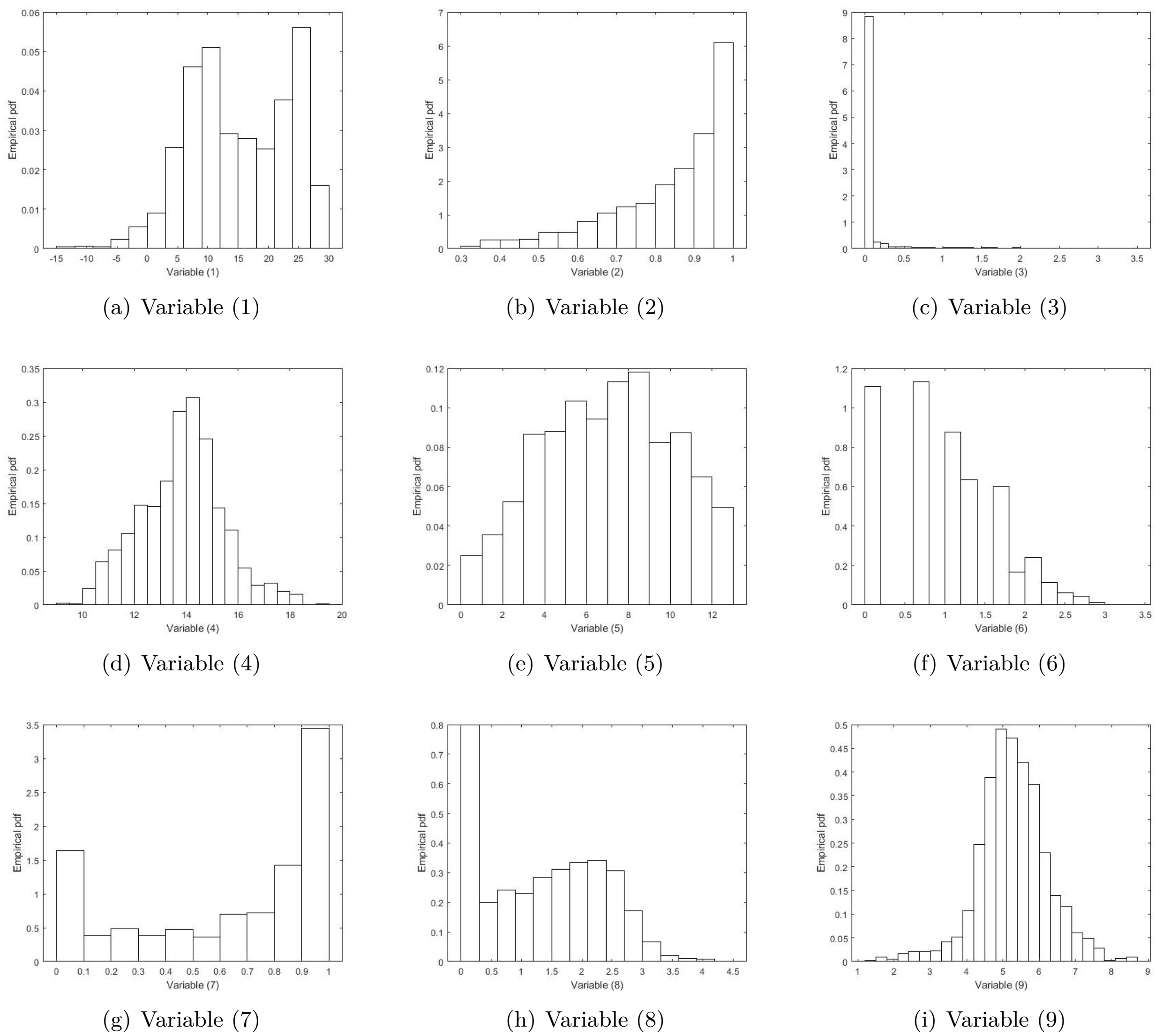
38. The original data used in Gennaioli et al. (2013) contains 1,569 subnational regions in 110 different countries, comprising 97% of the world's GDP.

39. Gennaioli et al. (2013) contains a large variety of variables. Yet, some of them, like the variable trust in others, cannot be used together with the primary data set unless one would accept a considerable shrinkage of sample size.

40. This, following Gennaioli et al. (2013), is the only prospective source among control variables that may cause the endogeneity issue in this data set.

carried out by the IV estimation method. γ is the coefficient for the added residual term discussed in Section 3.3.2 for Copula-IV and also represents the coefficient associated with the control variable in Qu and Lee (2015) (the δ parameter in Eq.(3.2)). Table 11 shows that the spatial spillover almost disappears when the country fixed-effects are incorporated since although the $\hat{\lambda}$ is significant using a traditional SAR model (SAR-IV), the estimates for the spatial coefficient are insignificant under both Qu-IV and Copula-IV. Similar to Yang et al. (2022), while our Copula-based estimator corrects the possible endogeneity through a significant γ and a set of different estimates from SAR-IV, the results may not be the same as those using the method with excluded IV (Qu-IV), since we estimate a distinct model. We do not know the actual data-generating process, so we cannot tell which estimator has the slightest bias. However, note that we reach a small discrepancy between Qu-IV and Copula-IV for the estimates of $\hat{\lambda}$.

In Table 12, we compare the estimates from SAR-IV to those from the Copula-IV, while the control-function-based method drops out due to the reason mentioned above. Although the differences in the parameters may not be compelling, Copula-IV still points to a significant $\hat{\gamma}$, indicating the probability of an endogenous W_n constructed by the endogenous variable - years of education. Therefore, our Copula method turns out to be a valid alternative to solve the endogeneity issue in a SAR model, especially when the excluded instruments are hard to obtain.



(a) Variable (1)

(b) Variable (2)

(c) Variable (3)

(d) Variable (4)

(e) Variable (5)

(f) Variable (6)

(g) Variable (7)

(h) Variable (8)

(i) Variable (9)

Figure 1: Distributions of variables.

	SAR - IV	Qu - IV	Copula-IV	SAR - IV	Qu - IV	Copula-IV
Intercept	2.6332***	2.3709***	3.1602***	-	-	-
	(0.2767)	(0.2761)	(0.3387)	-	-	-
(1) Temperature	-0.0044*	0.0139***	-0.0257***	-0.0117***	-0.0111***	-0.0209***
	(0.0025)	(0.0046)	(0.0064)	(0.0028)	(0.0030)	(0.0047)
(2) Inverse distance to coast	0.9975***	0.7298***	1.3920***	0.3458	0.3167***	0.5124***
	(0.1345)	(0.1427)	(0.1849)	(0.0979)	(0.1071)	(0.1146)
(3) ln Oil production per capita	0.2947***	0.2684***	0.3347***	0.1879***	0.1923***	0.2109***
	(0.0394)	(0.0409)	(0.0429)	(0.0223)	(0.0232)	(0.0234)
(4) ln Population	0.0332***	0.0138	0.0575***	0.0139	-0.0012	0.0242**
	(0.0090)	(0.0102)	(0.0120)	(0.0094)	(0.0114)	(0.0109)
(5) Years of Education	0.1735***	0.2486***	0.0981***	0.2578***	0.3426***	0.2406***
	(0.0148)	(0.0246)	(0.0238)	(0.0119)	(0.0303)	(0.0197)
(6) ln No. ethnic groups	-0.0500	0.0289	-0.1437***	-0.0369**	-0.0097	-0.0737***
	(0.0220)	(0.0290)	(0.0350)	(0.0148)	(0.0188)	(0.0237)
λ	0.4216***	0.4073***	0.3933***	0.2012***	0.0795	0.0715
	(0.0477)	0.0486	(0.0505)	(0.0651)	(0.0717)	(0.0716)
γ	-	-0.0857***	0.2663***	-	-0.0823***	0.0991**
	-	(0.0186)	(0.0727)	-	(0.0319)	(0.0502)
Country FE	N	N	N	Y	Y	Y
Sample sizes	1433	1433	1433	1433	1433	1433

Standard errors in parentheses.

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

Table 11: Estimation results for Setting 1.

	SAR - IV	Copula-IV	SAR - IV	Copula-IV
Intercept	2.0865***	2.5661***	-	-
	(0.2478)	(0.2905)	-	-
(1) Temperature	-0.0017	-0.0182***	-0.0129***	-0.0188***
	(0.0022)	(0.0048)	(0.0028)	(0.0043)
(2) Inverse distance to coast	0.7808***	1.0163***	0.3377***	0.3977***
	(0.1180)	(0.1409)	(0.0988)	(0.1040)
(3) ln Oil production per capita	0.2547***	0.2649***	0.1889***	0.1930***
	(0.0342)	(0.0349)	(0.0222)	(0.0222)
(4) ln Population	0.0330***	0.0502***	0.0045	0.0135
	(0.0087)	(0.0100)	(0.0100)	(0.0111)
(5) Years of Education	0.1362***	0.0558***	0.2422***	0.2117***
	(0.0123)	(0.0209)	(0.0125)	(0.0211)
(6) ln No. ethnic groups	-0.0039	-0.0380	-0.0131	-0.0279
	(0.0221)	(0.0230)	(0.0166)	(0.0184)
(7) Prob same language	0.2036***	0.3206***	0.1092***	0.1444**
	(0.0469)	(0.0587)	(0.0587)	(0.0618)
(8) ln power density	0.0386	0.1042***	0.0007	0.0245
	(0.0206)	(0.0269)	(0.0147)	(0.0197)
(9) ln travel time	-0.0070	-0.0221	-0.0388**	-0.0437***
	(0.0202)	(0.0199)	(0.0153)	(0.0154)
λ	0.5098***	0.4934***	0.2519***	0.2484***
	(0.0410)	(0.0427)	(0.0640)	(0.0640)
γ	-	0.2717***	-	0.1028**
	-	(0.0682)	-	(0.0571)
Country FE	N	N	Y	Y
Sample sizes	1433	1433	1433	1433

Standard errors in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Estimation results for Setting 2.

8 Conclusion

In this paper, we consider the specifications and estimations for three variants of a cross-sectional SAR model that might have possible endogeneity issues. In the first and second variants, i.e., a SAR model with an endogenous spatial weight matrix and the one with endogenous heterogeneity, the endogenous variables enter the SAR model in a nonlinear manner, while they are added as linear regressors in the third variant. Unlike the control function approach, we propose directly modelling the correlations among the endogenous variables, the error term in the SAR outcome equation, and the exogenous variables that are independent of the error term but might be correlated with the endogenous variables using a Gaussian copula. We propose 3-stage estimation methods (3SPML and 3SIV) and establish their consistency and asymptotic normality by employing the theory of asymptotic inference under near-epoch dependence. A Monte Carlo study is carried out to investigate the finite sample properties of our estimators and verify their robustness against different settings and misspecification. We apply our estimation methods to study the spatial spillovers in regional productivity and demonstrate the usefulness of our proposed Copula method in empirical applications. In future research, we plan to extend the three variants of a cross-sectional SAR model to spatial panel data settings and might also consider combining the copula endogeneity correction technique with IV methods to address the omitted variable problem in a control function approach, which is an interesting finding in our Monte Carlo experiments.

Appendix A. Expressions related to the statistics

A.1 Partitioned quadratic formulation

$$
\begin{aligned}
&(v_i^*, z_i^{*\prime}, x_i^{*\prime})\, \tilde{P}^{-1}\, (v_i^*, z_i^{*\prime}, x_i^{*\prime})' \\
&= \left(v_i^* - (\rho_{vz}', \mathbf{0}_{k\times 1}) \begin{pmatrix} P_z & P_{xz}' \\ P_{xz} & P_x \end{pmatrix}^{-1} \begin{pmatrix} \hat{z}_i^* \\ \hat{x}_i^* \end{pmatrix} \right)' \left(1 - (\rho_{vz}', \mathbf{0}_{k\times 1}) \begin{pmatrix} P_z & P_{xz}' \\ P_{xz} & P_x \end{pmatrix}^{-1} \begin{pmatrix} \rho_{vz} \\ \mathbf{0}_{k\times 1} \end{pmatrix} \right)^{-1} \\
&\quad \cdot \left(v_i^* - (\rho_{vz}', \mathbf{0}_{k\times 1}) \begin{pmatrix} P_z & P_{xz}' \\ P_{xz} & P_x \end{pmatrix}^{-1} \begin{pmatrix} \hat{z}_i^* \\ \hat{x}_i^* \end{pmatrix} \right) + (\hat{z}_i^{*\prime}, \hat{x}_i^{*\prime}) \begin{pmatrix} P_z & P_{xz}' \\ P_{xz} & P_x \end{pmatrix}^{-1} \begin{pmatrix} \hat{z}_i^* \\ \hat{x}_i^* \end{pmatrix} \\
&= \left(v_i^* - \rho_{vz}'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}(\hat{z}_i^* - P_{xz}'P_x^{-1}\hat{x}_i^*) \right)' \left(1 - \rho_{vz}'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}\rho_{vz} \right)^{-1} \\
&\quad \cdot \left(v_i^* - \rho_{vz}'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}(\hat{z}_i^* - P_{xz}'P_x^{-1}\hat{x}_i^*) \right) + (\hat{z}_i^{*\prime}, \hat{x}_i^{*\prime}) \begin{pmatrix} P_z & P_{xz}' \\ P_{xz} & P_x \end{pmatrix}^{-1} \begin{pmatrix} \hat{z}_i^* \\ \hat{x}_i^* \end{pmatrix} \\
&= \left(v_i^* - \rho_{vz}'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}(\hat{z}_i^* - P_{xz}'P_x^{-1}\hat{x}_i^*) \right)' \left(1 - \rho_{vz}'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}\rho_{vz} \right)^{-1} \\
&\quad \cdot \left(v_i^* - \rho_{vz}'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}(\hat{z}_i^* - P_{xz}'P_x^{-1}\hat{x}_i^*) \right) \\
&\quad + (\hat{z}_i^* - P_{xz}'P_x^{-1}\hat{x}_i^*)'(P_z - P_{xz}'P_x^{-1}P_{xz})^{-1}(\hat{z}_i^* - P_{xz}'P_x^{-1}\hat{x}_i^*) + \hat{x}_i^* P_x^{-1}\hat{x}_i \\
&= \frac{1}{\kappa} \left[v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - P_{xz}'P_x^{-1}x_i^*) \right]' \left[v_i^* - \rho_{vz}'\Xi^{-1}(z_i^* - P_{xz}'P_x^{-1}x_i^*) \right] \\
&\quad + (z_i^* - P_{xz}'P_x^{-1}x_i^*)'\Xi^{-1}(z_i^* - P_{xz}'P_x^{-1}x_i^*) + x_i^{*\prime}P_x^{-1}x_i^*
\end{aligned}
$$

where $\kappa = 1 - \rho_{vz}'\Xi^{-1}\rho_{vz}$, $\Xi = P_z - P_{xz}'P_x^{-1}P_{xz}$. The first and third equality holds by the partitioned quadratic formulation and the inverse of block matrix

$$
\begin{pmatrix} P_z & P_{xz}' \\ P_{xz} & P_x \end{pmatrix}^{-1} = \begin{pmatrix} (P_z - P_{xz}'P_xP_{xz})^{-1} & -(P_z - P_{xz}'P_xP_{xz})^{-1}P_{xz}'P_x^{-1} \\ -P_x^{-1}P_{xz}(P_z - P_{xz}'P_xP_{xz})^{-1} & P_x^{-1} + P_x^{-1}P_{xz}(P_z - P_{xz}'P_xP_{xz})^{-1}P_{xz}P_x^{-1} \end{pmatrix}^{-1}.
$$

A.2 Determinant of block matrix

$$
\begin{aligned}
\begin{vmatrix} 1 & \rho_{vz}' & \mathbf{0}_{k\times 1}' \\ \rho_{vz} & P_z & P_{xz}' \\ \mathbf{0}_{k\times 1} & P_{xz} & P_x \end{vmatrix} &= \left| \begin{pmatrix} P_z & P_{xz}' \\ P_{xz} & P_x \end{pmatrix} - \begin{pmatrix} \rho_{vz} \\ \mathbf{0}_{k\times 1} \end{pmatrix} \begin{pmatrix} \rho_{vz}', \mathbf{0}_{k\times 1}' \end{pmatrix} \right| \\
&= \begin{vmatrix} P_z - \rho_{vz}\rho_{vz}' & P_{xz}' \\ P_{xz} & P_z \end{vmatrix} \\
&= |P_z - \rho_{vz}\rho_{vz}'| \cdot \left| P_x - P_{xz}(P_z - \rho_{vz}\rho_{vz}')^{-1} P_{xz}' \right| \\
&= |P_z - \rho_{vz}\rho_{vz}'| \cdot |P_x| \cdot \left| I_k - P_x^{-1} P_{xz}(P_z - \rho_{vz}\rho_{vz}') P_{xz}' \right| \\
&= |P_x| \cdot \left| P_z - P_{xz}' P_x^{-1} P_{xz} - \rho_{vz}\rho_{vz}' \right| \\
&= |P_x| \cdot \left| P_z - P_{xz}' P_x^{-1} P_{xz} \right| \cdot \left| 1 - \rho_{vz}'(P_z - P_{xz}' P_x^{-1} P_{xz})^{-1} \rho_{vz} \right|,
\end{aligned}
$$

where the fourth and last equalities hold by applying the Sylvester's determinant theorem.

A.3 PMLE derivation

A.3.1 Expectation of the log pseudo-likelihood function

For a SAR model with endogenous W_n, note that $V_n(\omega) - (\mathscr{O}_n^{\perp} Z_n^*)\chi = S_n(\lambda)Y_n - X_n\beta - (\mathscr{O}_n^{\perp} Z_n^*)\chi = (\lambda_0 - \lambda)G_n X_n \beta_0 + X_n(\beta_0 - \beta) + (\mathscr{O}_n^{\perp} Z_n^*)(\chi_0 - \chi) + S_n(\lambda)S_n^{-1}V_n - (\mathscr{O}_n^{\perp} Z_n^*)\chi_0$, where $G_n = W_n S_n^{-1}$ with $S_n = I_n - \lambda_0 W_n$, thus

$$
\begin{aligned}
&\frac{1}{n}[V_n(\omega) - (\mathscr{O}_n^{\perp} Z_n^*)\chi]'[V_n(\omega) - (\mathscr{O}_n^{\perp} Z_n^*)\chi] \\
&\quad = \frac{1}{n} V_n' S_n'^{-1} S_n'(\lambda) S_n(\lambda) S_n^{-1} V_n - \frac{1}{n}\chi_0'(\mathscr{O}_n^{\perp} Z_n^*)'(\mathscr{O}_n^{\perp} Z_n^*)\chi_0 - \frac{2}{n}[S_n(\lambda)S_n^{-1}V_n - (\mathscr{O}_n^{\perp} Z_n^*)\chi_0]'(\mathscr{O}_n^{\perp} Z_n^*)\chi \\
&\qquad + (\lambda_0 - \lambda, \beta_0' - \beta', \chi_0' - \chi')\mathscr{A}_n(\lambda_0 - \lambda, \beta_0' - \beta', \chi_0' - \chi')'
\end{aligned}
$$

where $\mathscr{A}_n = \frac{1}{n}(G_n X_n \beta_0, X_n, \mathscr{O}_n^{\perp} Z_n^*)'(G_n X_n \beta_0, X_n, \mathscr{O}_n^{\perp} Z_n^*)$. By Lemma 2, $\frac{1}{n}\mathrm{E}(\ln L_n(\theta_{ML})) = \frac{1}{n}\mathrm{E}(\ln L_{n0}(\theta_{ML})) + o_p(1)$, the expectation of the log pseudo-likelihood function in Eq.(23) is

$$
\begin{aligned}
&\frac{1}{n}\mathrm{E}(\ln L_n(\theta_{ML})) \\
&\quad = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma_\xi^2 - \frac{1}{2}\ln|P_x| - \frac{1}{2}\ln|\Xi| + \frac{1}{n}\mathrm{E}(\ln|S_n(\lambda)|) - \frac{1}{2}\mathrm{tr}(P_x^{-1}P_{x,0}) - \frac{1}{2}\mathrm{tr}(\Xi^{-1}\Xi_0) \\
&\qquad - \frac{1}{2n}\sum_{i=1}^{n} x_i^{*\prime}(\Gamma_0 - \Gamma)\Xi^{-1}(\Gamma_0 - \Gamma)' x_i^* - \frac{1}{2\sigma_\xi^2}(\lambda_0 - \lambda, \beta_0' - \beta', \chi_0' - \chi')\mathrm{E}(\mathscr{A}_n)(\lambda_0 - \lambda, \beta_0' - \beta', \chi_0' - \chi')' \\
&\qquad - \frac{1}{2n}\frac{\sigma_{v,0}^2}{\sigma_\xi^2}\mathrm{E}[\mathrm{tr}(S_n'^{-1} S_n'(\lambda) S_n(\lambda) S_n^{-1})] + \frac{\sigma_{v,0}^2(1 - k_0)}{2\sigma_\xi^2} + \frac{1}{n}\frac{\sigma_{v,0}}{\sigma_\xi^2}\rho_{vz,0}'\chi\mathrm{E}[\mathrm{tr}(S_n(\lambda)S_n^{-1})] - \frac{\sigma_{v,0}}{\sigma_\xi^2}\rho_{vz,0}'\chi \\
&\qquad - \frac{1}{2}P_{x,0} + \frac{1}{2}P_{z,0} + o_p(1).
\end{aligned}
$$

For a SAR model with endogenous heterogeneity, as $V_n(\omega) - (\mathscr{O}_n^{\perp} Z_n^*)\chi = S_n(\zeta)Y_n - X_n\beta - (\mathscr{O}_n^{\perp} Z_n^*)\chi = [\Lambda(\zeta_0, Z_n) - \Lambda(\zeta, Z_n)]G_n X_n \beta_0 + X_n(\beta_0 - \beta) + (\mathscr{O}_n^{\perp} Z_n^*)(\chi_0 - \chi) + S_n(\zeta)S_n^{-1}V_n - (\mathscr{O}_n^{\perp} Z_n^*)\chi_0$, where $G_n = W_n S_n^{-1}$ with $S_n = I_n - \Lambda(\zeta_0, Z_n)W_n$, hence

$$
\begin{aligned}
&\frac{1}{n}[V_n(\omega) - (\mathscr{O}_n^{\perp} Z_n^*)\chi]'[V_n(\omega) - (\mathscr{O}_n^{\perp} Z_n^*)\chi] \\
&\quad = \frac{1}{n}V_n' S_n'^{-1} S_n'(\zeta) S_n(\zeta) S_n^{-1} V_n - \frac{1}{n}\chi_0'(\mathscr{O}_n^{\perp} Z_n^*)'(\mathscr{O}_n^{\perp} Z_n^*)\chi_0 - \frac{2}{n}[S_n(\zeta)S_n^{-1}V_n - (\mathscr{O}_n^{\perp} Z_n^*)\chi_0]'(\mathscr{O}_n^{\perp} Z_n^*)\chi \\
&\qquad + \frac{1}{n}(G_n X_n \beta_0)' \mathrm{diag}\{[\lambda(\zeta_0, z_1) - \lambda(\zeta, z_1)]^2, ..., [\lambda(\zeta_0, z_n) - \lambda(\zeta, z_n)]^2\}(G_n X_n \beta_0) \\
&\qquad + \frac{2}{n}(G_n X_n \beta_0)' \mathrm{diag}\{\lambda(\zeta_0, z_1) - \lambda(\zeta, z_1), ..., \lambda(\zeta_0, z_n) - \lambda(\zeta, z_n)\}(X_n, \mathscr{O}_n^{\perp} Z_n^*)(\beta_0' - \beta', \chi_0' - \chi')' \\
&\qquad + (\beta_0' - \beta', \chi_0' - \chi')[\frac{1}{n}(X_n, \mathscr{O}_n^{\perp} Z_n^*)'(X_n, \mathscr{O}_n^{\perp} Z_n^*)](\beta_0' - \beta', \chi_0' - \chi')'
\end{aligned}
$$

The expectation of the log pseudo-likelihood function is

$$
\begin{aligned}
&\frac{1}{n}\mathrm{E}(\ln L_n(\theta_{ML})) \\
&\quad = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma_\xi^2 - \frac{1}{2}\ln|P_x| - \frac{1}{2}\ln|\Xi| + \frac{1}{n}\mathrm{E}(\ln|S_n(\zeta)|) - \frac{1}{2}\mathrm{tr}(P_x^{-1}P_{x,0}) - \frac{1}{2}\mathrm{tr}(\Xi^{-1}\Xi_0) \\
&\qquad - \frac{1}{2n}\sum_{i=1}^{n} x_i^{*\prime}(\Gamma_0 - \Gamma)\Xi^{-1}(\Gamma_0 - \Gamma)' x_i^* - \frac{1}{2\sigma_\xi^2}(\beta_0' - \beta', \chi_0' - \chi')[\frac{1}{n}\mathrm{E}(X_n, \mathscr{O}_n^{\perp} Z_n^*)'(X_n, \mathscr{O}_n^{\perp} Z_n^*)](\beta_0' - \beta', \chi_0' - \chi')' \\
&\qquad - \frac{1}{2n}\frac{\sigma_{v,0}^2}{\sigma_\xi^2}\mathrm{E}[\mathrm{tr}(S_n'^{-1} S_n'(\zeta) S_n(\zeta) S_n^{-1})] + \frac{\sigma_{v,0}^2(1-k_0)}{2\sigma_\xi^2} + \frac{1}{n}\frac{\sigma_{v,0}}{\sigma_\xi^2}\rho_{vz,0}'\chi \mathrm{E}[\mathrm{tr}(S_n(\zeta)S_n^{-1})] - \frac{\sigma_{v,0}}{\sigma_\xi^2}\rho_{vz,0}'\chi \\
&\qquad - \frac{1}{n}\frac{1}{2\sigma_\xi^2}\mathrm{E}\left[(G_n X_n \beta_0)' \mathrm{diag}\{[\lambda(\zeta_0, z_1) - \lambda(\zeta, z_1)]^2, ..., [\lambda(\zeta_0, z_n) - \lambda(\zeta, z_n)]^2\}(G_n X_n \beta_0)\right] \\
&\qquad - \frac{1}{n}\frac{1}{\sigma_\xi^2}\mathrm{E}[(G_n X_n \beta_0)' \mathrm{diag}\{\lambda(\zeta_0, z_1) - \lambda(\zeta, z_1), ..., \lambda(\zeta_0, z_n) - \lambda(\zeta, z_n)\}(X_n, \mathscr{O}_n^{\perp} Z_n^*)(\beta_0' - \beta', \chi_0' - \chi')'] \\
&\qquad - \frac{1}{2}P_{x,0} + \frac{1}{2}P_{z,0} + o_p(1).
\end{aligned}
$$

A.3.2 First and second order derivatives

For a SAR model with endogenous spatial weights, the first order derivatives are

$$
\begin{aligned}
\frac{\partial \ln L_n(\theta_{ML})}{\partial \lambda} &= \frac{1}{\sigma_\xi^2}(W_n Y_n)'\left[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right] - \mathrm{tr}[W_n S_n^{-1}(\lambda)]; \\
\frac{\partial \ln L_n(\theta_{ML})}{\partial \beta} &= \frac{1}{\sigma_\xi^2}X_n'\left[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right]; \\
\frac{\partial \ln L_n(\theta_{ML})}{\partial \sigma_\xi^2} &= -\frac{n}{2\sigma_\xi^2} + \frac{1}{2\sigma_\xi^4}\left[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right]'\left[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right]; \\
\frac{\partial \ln L_n(\theta_{ML})}{\partial \chi} &= \frac{1}{\sigma_\xi^2}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'\left[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right]; \\
\frac{\partial \ln L_n(\theta_{ML})}{\partial \alpha} &= -\frac{n}{2}\frac{\partial \ln|P_x|}{\partial \alpha} - \frac{1}{2}\frac{\partial}{\partial \alpha}\mathrm{tr}\left[P_x^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*\right]; \\
\frac{\partial \ln L_n(\theta_{ML})}{\partial \delta} &= -\frac{n}{2}\frac{\partial \ln|\Xi|}{\partial \delta} - \frac{1}{2}\frac{\partial}{\partial \delta}\mathrm{tr}\left[\Xi^{-1}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\right],
\end{aligned}
$$

where $V_n(\omega) = S_n(\lambda)Y_n - X_n\beta$, $S_n(\lambda) = I_n - \lambda W_n$. α is a J_1-dimensional column vector of distinct elements in P_x, the J_1-dimensional vector $\frac{\partial \ln|P_x|}{\partial \alpha}$ has the j_1th element $\mathrm{tr}(P_x^{-1}\frac{\partial P_x}{\partial \alpha_{j_1}})$ and $\frac{\partial}{\partial \alpha}\mathrm{tr}[P_x^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*]$

has its j_1th element $-\mathrm{tr}(P_x^{-1}\frac{\partial P_x}{\partial \alpha_{j_1}}P_x^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*)$ for $j_1 = 1, ..., J_1$. δ is a J_2-dimensional column vector of distinct elements in Ξ, the J_2-dimensional vector $\frac{\partial \ln|\Xi|}{\partial \delta}$ has the j_2th element $\mathrm{tr}(\Xi^{-1}\frac{\partial \Xi}{\partial \delta_{j_2}})$ and $\frac{\partial}{\partial \delta}\mathrm{tr}[\Xi^{-1}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)]$ has its j_2th element $-\mathrm{tr}[\Xi^{-1}\frac{\partial \Xi}{\partial \delta_{j_2}}\Xi^{-1}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)]$ for $j_2 = 1, ..., J_2$. 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 $\frac{\partial \ln L_n(\theta_{ML})}{\partial \lambda}$ should be replaced by

$$\frac{\partial \ln L_n(\theta_{ML})}{\partial \zeta_\iota} = \frac{1}{\sigma_\xi^2}\left[\mathrm{diag}\left(\frac{\partial \lambda(\zeta, z_i)}{\partial \zeta_\iota}\right)W_nY_n\right]'\left[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right] - \mathrm{tr}\left[\mathrm{diag}\left(\frac{\partial \lambda(\zeta, z_i)}{\partial \zeta_\iota}\right)W_nS_n^{-1}(\zeta)\right],\ \iota = 1, ..., p_0.$$

The second order derivatives for the endogenous W_n specification are

$$\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \lambda \partial \lambda} = -\frac{1}{\sigma_\xi^2}(W_nY_n)'(W_nY_n) - \mathrm{tr}[W_nS_n^{-1}(\lambda)]^2;\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \lambda \partial \beta} = -\frac{1}{\sigma_\xi^2}(W_nY_n)'X_n;$$
$$\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \lambda \partial \sigma_\xi^2} = -\frac{1}{\sigma_\xi^4}(W_nY_n)'\left[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right];\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \lambda \partial \chi} = -\frac{1}{\sigma_\xi^2}(W_nY_n)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*);$$
$$\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \lambda \partial \alpha} = 0;\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \lambda \partial \delta} = 0;$$
$$\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \beta \partial \beta'} = -\frac{1}{\sigma_\xi^2}X_n'X_n;\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \beta \partial \sigma_\xi^2} = -\frac{1}{\sigma_\xi^4}X_n'\left[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right];$$
$$\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \beta \partial \chi'} = -\frac{1}{\sigma_\xi^2}X_n'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*);\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \beta \partial \alpha'} = 0;\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \beta \partial \delta'} = 0;$$
$$\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \sigma_\xi^2 \partial \sigma_\xi^2} = \frac{n}{2\sigma_\xi^4} - \frac{1}{\sigma_\xi^6}\left[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right]'\left[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right];$$
$$\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \sigma_\xi^2 \partial \chi} = -\frac{1}{\sigma_\xi^4}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'\left[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right];\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \sigma_\xi^2 \partial \alpha} = 0;\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \sigma_\xi^2 \partial \delta} = 0;$$
$$\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \chi \partial \chi'} = -\frac{1}{\sigma_\xi^2}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*);\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \chi \partial \alpha'} = 0;\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \chi \partial \delta'} = 0;$$
$$\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \alpha \partial \alpha'} = -\frac{n}{2}\frac{\partial^2 \ln|P_x|}{\partial \alpha \partial \alpha'} - \frac{1}{2}\frac{\partial^2}{\partial \alpha \partial \alpha'}\mathrm{tr}[P_x^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*];\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \alpha \partial \delta'} = 0;$$
$$\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial \delta \partial \delta'} = -\frac{n}{2}\frac{\partial^2 \ln|\Xi|}{\partial \delta \partial \delta'} - \frac{1}{2}\frac{\partial^2}{\partial \delta \partial \delta'}\mathrm{tr}[\Xi^{-1}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)],$$

where $\frac{\partial^2 \ln|P_x|}{\partial \alpha \partial \alpha'}$ is a $J_1 \times J_1$ matrix with the (j_1, k_1)th element $\frac{\partial^2 \ln|P_x|}{\partial \alpha_{j_1}\partial \alpha_{k_1}} = -\mathrm{tr}(P_x^{-1}\frac{\partial P_x}{\partial \alpha_{k_1}}P_x^{-1}\frac{\partial P_x}{\partial \alpha_{j_1}})$ and the (j_1, k_1)th element of $\frac{\partial^2}{\partial \alpha \partial \alpha'}\mathrm{tr}\left[P_x^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*\right]$ is

$$\frac{\partial^2}{\partial \alpha_{j_1}\partial \alpha_{k_1}}\mathrm{tr}\left[P_x^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*\right]$$
$$= \mathrm{tr}\left(P_x^{-1}\left(\frac{\partial P_x}{\partial \alpha_{k_1}}P_x^{-1}\frac{\partial P_x}{\partial \alpha_{j_1}} + \frac{\partial P_x}{\partial \alpha_{j_1}}P_x^{-1}\frac{\partial P_x}{\partial \alpha_{k_1}}\right)P_x^{-1}\hat{X}_n^{*\prime}\hat{X}_n^*\right)$$

for $j_1, k_1 = 1, ..., J_1$. $\frac{\partial^2 \ln|\Xi|}{\partial \delta \partial \delta'}$ is a $J_2 \times J_2$ matrix with the (j_2, k_2)th element $\frac{\partial^2 \ln|\Xi|}{\partial \delta_{j_2}\partial \delta_{k_2}} = -\mathrm{tr}(\Xi^{-1}\frac{\partial \Xi}{\partial \delta_{k_2}}\Xi^{-1}\frac{\partial \Xi}{\partial \delta_{j_2}})$ and the (j_2, k_2)th element of $\frac{\partial^2}{\partial \delta \partial \delta'}\mathrm{tr}\left[\Xi^{-1}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\right]$ is

$$\frac{\partial^2}{\partial\delta_{j_2}\partial\delta_{k_2}}\operatorname{tr}\left[\Xi^{-1}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\right]$$
$$=\operatorname{tr}\left(\Xi^{-1}\left(\frac{\partial\Xi}{\partial\delta_{k_2}}\Xi^{-1}\frac{\partial\Xi}{\partial\delta_{j_2}}+\frac{\partial\Xi}{\partial\delta_{j_2}}\Xi^{-1}\frac{\partial\Xi}{\partial\delta_{k_2}}\right)\Xi^{-1}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\right)$$

for $j_2, k_2 = 1, ..., J_2$. For the endogenous heterogeneity specification, the differences are that $V_n(\omega) = S_n(\zeta)Y_n - X_n\beta$, and that the first six derivatives associated with partial derivatives with respect to λ should be replaced by

$$\begin{aligned}
\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\zeta_\iota\partial\zeta_\iota} &= -\frac{1}{\sigma_\xi^2}\left[\operatorname{diag}\left(\frac{\partial\lambda(\zeta,z_i)}{\partial\zeta_\iota}\right)W_nY_n\right]'\left[\operatorname{diag}\left(\frac{\partial\lambda(\zeta,z_i)}{\partial\zeta_\iota}\right)W_nY_n\right]-\operatorname{tr}\left[\operatorname{diag}\left(\frac{\partial^2\lambda(\zeta,z_i)}{\partial\zeta_\iota\partial\zeta_\iota}\right)W_nS_n^{-1}(\zeta)\right]\\
&\quad-\operatorname{tr}\left[\operatorname{diag}\left(\frac{\partial\lambda(\zeta,z_i)}{\partial\zeta_\iota}\right)W_nS_n^{-1}(\zeta)\right]^2, \iota = 1,\dots,p_0;\\
\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\zeta_{\iota_1}\partial\zeta_{\iota_2}} &= -\frac{1}{\sigma_\xi^2}\left[\operatorname{diag}\left(\frac{\partial\lambda(\zeta,z_i)}{\partial\zeta_{\iota_1}}\right)W_nY_n\right]'\left[\operatorname{diag}\left(\frac{\partial\lambda(\zeta,z_i)}{\partial\zeta_{\iota_2}}\right)W_nY_n\right]-\operatorname{tr}\left[\operatorname{diag}\left(\frac{\partial^2\lambda(\zeta,z_i)}{\partial\zeta_{\iota_1}\partial\zeta_{\iota_2}}\right)W_nS_n^{-1}(\zeta)\right]\\
&\quad-\operatorname{tr}\left[\operatorname{diag}\left(\frac{\partial\lambda(\zeta,z_i)}{\partial\zeta_{\iota_1}}\cdot\frac{\partial\lambda(\zeta,z_i)}{\partial\zeta_{\iota_2}}\right)\left(W_nS_n^{-1}(\zeta)\right)^2\right], \iota_1\neq\iota_2 \text{ and } \iota_1,\iota_2 = 1,\dots,p_0;\\
\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\zeta_\iota\partial\beta} &= -\frac{1}{\sigma_\xi^2}X_n'\left[\operatorname{diag}\left(\frac{\partial\lambda(\zeta,z_i)}{\partial\zeta_\iota}\right)W_nY_n\right];\\
\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\zeta_\iota\partial\sigma_\xi^2} &= -\frac{1}{\sigma_\xi^4}\left[\operatorname{diag}\left(\frac{\partial\lambda(\zeta,z_i)}{\partial\zeta_\iota}\right)W_nY_n\right]'\left[V_n(\omega)-(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi\right];\\
\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\zeta_\iota\partial\chi} &= -\frac{1}{\sigma_\xi^2}\left[\operatorname{diag}\left(\frac{\partial\lambda(\zeta,z_i)}{\partial\zeta_\iota}\right)W_nY_n\right]'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*);\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\zeta_\iota\partial\alpha}=0;\ \frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\zeta_\iota\partial\delta}=0.
\end{aligned}$$

A.3.3 The variance-covariance matrix

The variance-covariance matrix $\mathscr{I}_{ML,0}^{-1} = -\left(\lim_{n\to\infty}\frac{1}{n}\mathrm{E}\left(\frac{\partial^2\ln L_n(\theta_{ML,0})}{\partial\theta_{ML}\partial\theta_{ML}'}\right)\right)^{-1}$, where $\mathscr{I}_{ML,0}$ is the Fisher information matrix. For a SAR model with endogenous W_n,

$$\begin{aligned}
&\mathrm{E}\left(\frac{\partial^2\ln L_n(\theta_{ML,0})}{\partial\theta_{ML}\partial\theta_{ML}'}\right)\\
&= -\frac{1}{\sigma_{\xi,0}^2}\begin{bmatrix}
A_{\lambda\lambda} & \mathrm{E}(G_nX_n\beta_0)'X_n & \operatorname{tr}[\mathrm{E}(G_n)] & \sigma_{v,0}\rho_{vz,0}'\operatorname{tr}[\mathrm{E}(G_n)] & 0 & 0\\
* & X_n'X_n & 0 & 0 & 0 & 0\\
* & * & \frac{n}{2\sigma_{\xi,0}^2} & 0 & 0 & 0\\
* & * & * & n\Xi_0 & 0 & 0\\
* & * & * & * & A_{\alpha\alpha} & 0\\
* & * & * & * & * & A_{\delta\delta}
\end{bmatrix} \qquad \text{(A.1)}
\end{aligned}$$

where $A_{\lambda\lambda} = \mathrm{E}[(G_nX_n\beta_0)'(G_nX_n\beta_0)]+\operatorname{tr}[\mathrm{E}(\sigma_{\xi,0}^2G_n^2+\sigma_{v,0}^2G_n'G_n)]$; $(A_{\alpha\alpha})_{k_1j_1} = \frac{n\sigma_{\xi,0}^2}{2}\operatorname{tr}(P_{x,0}^{-1}\frac{\partial P_{x,0}}{\partial\alpha_{k_1}}P_{x,0}^{-1}\frac{\partial P_{x,0}}{\partial\alpha_{j_1}})$ for $j_1, k_1 = 1, ..., J_1$; $(A_{\delta\delta})_{k_2j_2} = \frac{n\sigma_{\xi,0}^2}{2}\operatorname{tr}(\Xi_0^{-1}\frac{\partial\Xi_0}{\partial\delta_{k_2}}\Xi_0^{-1}\frac{\partial\Xi_0}{\partial\delta_{j_2}})$ for $j_2, k_2 = 1, ..., J_2$. For a SAR model with endogenous heterogeneity,

$$\mathrm{E}\left(\frac{\partial^2 \ln L_n(\theta_{ML,0})}{\partial\theta_{ML}\partial\theta'_{ML}}\right)$$
$$= -\frac{1}{\sigma^2_{\xi,0}}\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p_0} & \mathrm{E}(\widetilde{G}_{n,1}X_n\beta_0)'X_n & A_{\zeta_1,\sigma^2_\xi} & A_{\zeta_1,\chi} & 0 & 0 \\ * & a_{22} & \dots & a_{2p_0} & \mathrm{E}(\widetilde{G}_{n,2}X_n\beta_0)'X_n & A_{\zeta_2,\sigma^2_\xi} & A_{\zeta_2,\chi} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ * & * & \dots & a_{p_0p_0} & \mathrm{E}(\widetilde{G}_{n,p_0}X_n\beta_0)'X_n & A_{\zeta_{p_0},\sigma^2_\xi} & A_{\zeta_{p_0},\chi} & 0 & 0 \\ * & * & \dots & * & X'_nX_n & 0 & 0 & 0 & 0 \\ * & * & \dots & * & * & A_{\sigma^2_\xi\sigma^2_\xi} & 0 & 0 & 0 \\ * & * & \dots & * & * & * & n\Xi_0 & 0 & 0 \\ * & * & \dots & * & * & * & * & A_{\alpha\alpha} & 0 \\ * & * & \dots & * & * & * & * & * & A_{\delta\delta} \end{bmatrix} \tag{A.2}$$

where $G_n = W_nS_n^{-1}(\zeta_0)$, $\tilde{G}_{n,\iota} = \mathrm{diag}(\frac{\partial\lambda(\zeta,z_i)}{\partial\zeta_\iota})G_n$, $\tilde{G}_{n,\iota\iota} = \mathrm{diag}\left(\frac{\partial^2\lambda(\zeta,z_i)}{\partial\zeta_\iota\partial\zeta_\iota}\right)G_n$, $\tilde{G}_{n,\iota_1\iota_2} = \mathrm{diag}\left(\frac{\partial^2\lambda(\zeta,z_i)}{\partial\zeta_{\iota_1}\partial\zeta_{\iota_2}}\right)G_n$; $a_{\iota\iota} = \mathrm{E}[(\tilde{G}_{n,\iota}X_n\beta_0)'(\tilde{G}_{n,\iota}X_n\beta_0)] + \sigma^2_{v,0}\mathrm{tr}[\mathrm{E}(\tilde{G}'_{n,\iota}\tilde{G}_{n,\iota})] + \sigma^2_{\xi,0}\mathrm{tr}[\mathrm{E}(\tilde{G}^2_{n,\iota} + \tilde{G}_{n,\iota\iota})]$, $\iota = 1,...,p_0$; $a_{\iota_1\iota_2} = \mathrm{E}[(\tilde{G}_{n,\iota_1}X_n\beta_0)'(\tilde{G}_{n,\iota_2}X_n\beta_0)] + \sigma^2_{v,0}\mathrm{tr}[\mathrm{E}(\tilde{G}'_{n,\iota_1}\tilde{G}_{n,\iota_2})] + \sigma^2_{\xi,0}\mathrm{tr}[\mathrm{E}(\tilde{G}_{n,\iota_1\iota_2} + \tilde{G}'_{n,\iota_1}\tilde{G}_{n,\iota_2})]$, $\iota_1 \neq \iota_2, \iota_1, \iota_2 = 1,...,p_0$; $A_{\zeta_\iota,\sigma^2_\xi} = \mathrm{tr}[\mathrm{E}(\tilde{G}_{n,\iota})]$, $\iota = 1,...,p_0$; $A_{\zeta_\iota,\chi} = \sigma_{v,0}\rho'_{vz,0}\,\mathrm{tr}[\mathrm{E}(\tilde{G}_{n,\iota})]$, $\iota = 1,...,p_0$; $A_{\sigma^2_\xi\sigma^2_\xi} = \frac{n}{2\sigma^2_{\xi,0}}$; $(A_{\alpha\alpha})_{k_1j_1} = \frac{n\sigma^2_{\xi,0}}{2}\mathrm{tr}(P^{-1}_{x,0}\frac{\partial P_{x,0}}{\partial\alpha_{k_1}}P^{-1}_{x,0}\frac{\partial P_{x,0}}{\partial\alpha_{j_1}})$ for $j_1, k_1 = 1,...,J_1$; $(A_{\delta\delta})_{k_2j_2} = \frac{n\sigma^2_{\xi,0}}{2}\mathrm{tr}(\Xi_0^{-1}\frac{\partial\Xi_0}{\partial\delta_{k_2}}\Xi_0^{-1}\frac{\partial\Xi_0}{\partial\delta_{j_2}})$ for $j_2, k_2 = 1,...,J_2$.

Appendix B. Mathematical proofs

Proof of Lemma 1. First, we prove part (i). $\frac{1}{n}\|X^*_n\|^2 = \frac{1}{n}\sum_{i=1}^n\|x^*_i\|^2 \leq \sup_i\|x^*_i\|^2 = \wp_0^2(k)$, the other three results can be shown in a similar way. Second, we consider part (ii). $\frac{1}{n}\|\hat{X}^*_n - X^*_n\|^2 = \frac{1}{n}\sum_{i=1}^n\sum_{\tau=1}^k\|\hat{x}^*_{\tau,i} - x^*_{\tau,i}\|^2 \leq \frac{1}{n}\sum_{i=1}^n\sum_{\tau=1}^k\frac{\log\log n}{n^{1-a_\tau}} = \log\log n\cdot(\sum_{\tau=1}^k n^{a_\tau-1})$. Similarly, we have $\frac{1}{n}\|\hat{Z}^*_n - Z^*_n\|^2 \leq \log\log n\cdot(\sum_{\iota=1}^k n^{a_\iota-1})$.

Proof of Proposition 2.

$$\begin{aligned}
&\frac{1}{n}\left[a'\varphi'_n(\theta)(\hat{\mathscr{O}}^\perp_n\hat{Z}^*_n)b - a'\varphi'_n(\theta)(\mathscr{O}^\perp_n Z^*_n)b\right] \\
&= \frac{1}{n}a'\varphi'_n(\theta)\hat{\mathscr{O}}^\perp_n(\hat{Z}^*_n - Z^*_n)b - \frac{1}{n}a'\varphi'_n(\theta)(\hat{\mathscr{O}}_n - \mathscr{O}_n)Z^*_nb \\
&= \frac{a'\varphi'_n(\theta)(\hat{Z}^*_n - Z^*_n)b}{n} - \frac{a'\varphi'_n(\theta)\hat{X}^*_n}{n}\left(\frac{\hat{X}^{*\prime}_n\hat{X}^*_n}{n}\right)^{-1}\frac{\hat{X}^{*\prime}_n(\hat{Z}^*_n - Z^*_n)b}{n} - \frac{a'\varphi'_n(\theta)(\hat{X}^*_n - X^*_n)}{n}\left(\frac{\hat{X}^{*\prime}_n\hat{X}^*_n}{n}\right)^{-1}\frac{\hat{X}^{*\prime}_nZ^*_nb}{n} \\
&\quad - \frac{a'\varphi'_n(\theta)X^*_n}{n}\left(\frac{X^{*\prime}_nX^*_n}{n}\right)^{-1}\frac{(\hat{X}^*_n - X^*_n)'Z^*_nb}{n} - \frac{a'\varphi'_n(\theta)X^*_n}{n}\left\{\left(\frac{\hat{X}^{*\prime}_n\hat{X}^*_n}{n}\right)^{-1} - \left(\frac{X^{*\prime}_nX^*_n}{n}\right)^{-1}\right\}\frac{\hat{X}^{*\prime}_nZ^*_nb}{n} \\
&= \mathscr{I}_1 - \mathscr{I}_2 - \mathscr{I}_3 - \mathscr{I}_4 - \mathscr{I}_5.
\end{aligned}$$

For any $m\times n$ matrix A, denote $\|A\|_o = [\mu_{\max}(A'A)]^{1/2}$ as the operator norm, where $\mu_{\max}$ is the largest eigenvalue of A. Note that $\|A\|_o \leq \|A\|$ by the spectral radius theorem. The desired result $\frac{1}{n}[a'\varphi'_n(\theta)(\hat{\mathscr{O}}^\perp_n\hat{Z}^*_n)b - a'\varphi'_n(\theta)(\mathscr{O}^\perp_nZ^*_n)b] = o_p(1)$ follows by triangular inequality if we show

$$|\mathscr{I}_1|, |\mathscr{I}_2|, |\mathscr{I}_3|, |\mathscr{I}_4|, |\mathscr{I}_5| = o_p(1).$$

For term $\mathscr{I}_1$,

$$|\mathscr{I}_1| \leq \frac{[a'\varphi_n'(\theta)\varphi_n(\theta)a]^{\frac{1}{2}} \cdot \left[b'(\hat{Z}_n^* - Z_n^*)'(\hat{Z}_n^* - Z_n^*)b\right]^{\frac{1}{2}}}{n} \leq \frac{[a'\varphi_n'(\theta)\varphi_n(\theta)a]^{\frac{1}{2}} \cdot (b'b)^{\frac{1}{2}} \cdot \left[\mu_{max}(\hat{Z}_n^* - Z_n^*)'(\hat{Z}_n^* - Z_n^*)\right]^{\frac{1}{2}}}{n}$$
$$= \frac{[a'\varphi_n'(\theta)\varphi_n(\theta)a]^{\frac{1}{2}} \cdot (b'b)^{\frac{1}{2}} \cdot \left\|\hat{Z}_n^* - Z_n^*\right\|_o}{n} \leq \frac{[a'\varphi_n'(\theta)\varphi_n(\theta)a]^{\frac{1}{2}} \cdot (b'b)^{\frac{1}{2}}}{\sqrt{n}} \cdot \left\|\frac{\hat{Z}_n^* - Z_n^*}{\sqrt{n}}\right\| = O_p(1)o_p(1) = o_p(1),$$

where the first inequality is from the Cauchy-Schwarz inequality, the second inequality holds as $(\hat{Z}_n^* - Z_n^*)'(\hat{Z}_n^* - Z_n^*)$ is non-negative definite, the last inequality holds because $\|A\|_o \leq \|A\|$, the penultimate equality holds as $\frac{[a'\varphi_n'(\theta)\varphi_n(\theta)a]\cdot(b'b)}{n} = O_p(1)$ and by Lemma 1(ii).

For term $\mathscr{I}_2$,

$$|\mathscr{I}_2| \leq \frac{\left[a'\varphi_n'(\theta)\hat{X}_n^*\hat{X}_n^{*\prime}\varphi_n(\theta)a\right]^{\frac{1}{2}}}{n} \left\|\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\right\|_o \frac{\left[b'(\hat{Z}_n^* - Z_n^*)'\hat{X}_n^*\hat{X}_n^{*\prime}(\hat{Z}_n^* - Z_n^*)b\right]^{\frac{1}{2}}}{n}$$
$$\leq \frac{[a'\varphi_n'(\theta)\varphi_n(\theta)a]^{\frac{1}{2}}\,\|\hat{X}_n^*\|_o}{n} \left\|\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\right\|_o \frac{(b'b)^{\frac{1}{2}}\|\hat{X}_n^*\|_o\|\hat{Z}_n^* - Z_n^*\|_o}{n}$$
$$\leq \frac{[a'\varphi_n'(\theta)\varphi_n(\theta)a]^{\frac{1}{2}} \cdot (b'b)^{\frac{1}{2}}}{\sqrt{n}} \cdot \left\|\frac{\hat{X}_n^*}{\sqrt{n}}\right\|^2 \left\|\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\right\|_o \cdot \left\|\frac{\hat{Z}_n^* - Z_n^*}{\sqrt{n}}\right\| = O_p(1)\tilde{C}_0\hat{\wp}_0^2(k)O_p(1)o_p(1) = o_p(1),$$

where the first and second inequalities hold by Cauchy-Schwarz inequality and that $\hat{X}_n^*\hat{X}_n^{*\prime}$ and $(\hat{Z}_n^* - Z_n^*)'(\hat{Z}_n^* - Z_n^*)$ are non-negative definite, the third inequality holds because $\|A\|_o \leq \|A\|$, the penultimate equality holds by Lemma 1. The desired results of term $|\mathscr{I}_3|$ and term $|\mathscr{I}_4|$ follow by similar argument used for term $|\mathscr{I}_2|$.

For term $\mathscr{I}_5$, note that

$$\mathscr{I}_5 = \frac{a'\varphi_n'(\theta)X_n^*}{n}\left(\frac{X_n^{*\prime}X_n^*}{n}\right)^{-1}\left\{\left(\frac{X_n^{*\prime}X_n^*}{n}\right) - \left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)\right\}\left(\frac{X_n^{*\prime}X_n^*}{n}\right)^{-1}\frac{\hat{X}_n^{*\prime}Z_n^*b}{n}$$
$$= -\frac{a'\varphi_n'(\theta)X_n^*}{n}\left(\frac{X_n^{*\prime}X_n^*}{n}\right)^{-1}\left(\frac{X_n^{*\prime}(\hat{X}_n^* - X_n^*)}{n}\right)\left(\frac{X_n^{*\prime}X_n^*}{n}\right)^{-1}\frac{\hat{X}_n^{*\prime}Z_n^*b}{n}$$
$$-\frac{a'\varphi_n'(\theta)X_n^*}{n}\left(\frac{X_n^{*\prime}X_n^*}{n}\right)^{-1}\left(\frac{(\hat{X}_n^* - X_n^*)'\hat{X}_n^*}{n}\right)\left(\frac{X_n^{*\prime}X_n^*}{n}\right)^{-1}\frac{\hat{X}_n^{*\prime}Z_n^*b}{n}$$

Then,

$$|\mathscr{I}_5| \leq O_p(1)[\tilde{C}_0\wp_0^2(k)]O_p(1)o_p(1)O_p(1)[\tilde{C}_0\hat{\wp}_0^2(k)]^{\frac{1}{2}}[\tilde{C}_0\wp_0^2(p)]^{\frac{1}{2}}$$
$$+ O_p(1)[\tilde{C}_0\wp_0^2(k)]^{\frac{1}{2}}O_p(1)o_p(1)O_p(1)[\tilde{C}_0\hat{\wp}_0^2(k)][\tilde{C}_0\wp_0^2(p)]^{\frac{1}{2}}$$
$$= o_p(1).$$

This finishes the proof.

Proof of Lemma 2. $\frac{1}{n}\left[\ln L_n\left(\theta_{ML}\right) - \ln L_{n0}\left(\theta_{ML}\right)\right]$ has three main components. The first component $\mathscr{A}_1$ is

$$\begin{aligned}
\mathscr{A}_1 &= -\frac{1}{n}\left([V_n(\omega)-(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi]'[V_n(\omega)-(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi]-[V_n(\omega)-(\mathscr{O}_n^{\perp}Z_n^*)\chi]'[V_n(\omega)-(\mathscr{O}_n^{\perp}Z_n^*)\chi]\right)\\
&= \frac{2}{n}V_n'(\omega)(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*-\mathscr{O}_n^{\perp}Z_n^*)\chi-\frac{1}{n}\left([(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi]'[(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)\chi]-[(\mathscr{O}_n^{\perp}Z_n^*)\chi]'[(\mathscr{O}_n^{\perp}Z_n^*)\chi]\right)\\
&= \frac{2}{n}V_n'(\omega)(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*-\mathscr{O}_n^{\perp}Z_n^*)\chi-\frac{1}{n}\chi'(\hat{Z}_n^{*\prime}\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*-Z_n^{*\prime}\mathscr{O}_n^{\perp}Z_n^*)\chi\\
&= \frac{2}{n}V_n'(\omega)(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*-\mathscr{O}_n^{\perp}Z_n^*)\chi-\frac{1}{n}\chi'(\hat{Z}_n^*-Z_n^*)'\hat{Z}_n^*\chi-\frac{1}{n}\chi'Z_n^{*\prime}(\hat{Z}_n^*-Z_n^*)\chi\\
&\quad+\frac{1}{n}\chi'(\hat{Z}_n^*-Z_n^*)'\hat{\mathscr{O}}_n\hat{Z}_n^*\chi+\frac{1}{n}\chi'Z_n^{*\prime}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*-\mathscr{O}_n^{\perp}Z_n^*)\chi\\
&= \mathscr{A}_{11}-\mathscr{A}_{12}-\mathscr{A}_{13}+\mathscr{A}_{14}+\mathscr{A}_{15}
\end{aligned}$$

where $V_n(\omega)=(\lambda_0-\lambda)G_nX_n\beta_0+X_n(\beta_0-\beta)+S_n(\lambda)S_n^{-1}V_n$ for endogenous spatial weights specification, and $V_n(\omega)=[\Lambda(\zeta_0,Z_n)-\Lambda(\zeta,Z_n)]W_nS_n^{-1}X_n\beta_0+X_n(\beta_0-\beta)+S_n(\zeta)S_n^{-1}V_n$ for endogenous heterogeneity specification. $|\mathscr{A}_{11}|,|\mathscr{A}_{15}|=o_p(1)$ by Proposition 2. For term $\mathscr{A}_{12}$,

$$|\mathscr{A}_{12}|\leq\frac{[\chi'(\hat{Z}_n^*-Z_n^*)'(\hat{Z}_n^*-Z_n^*)\chi]^{\frac{1}{2}}\cdot[\chi'\hat{Z}_n^{*\prime}\hat{Z}_n^*\chi]^{\frac{1}{2}}}{n}\leq(\chi'\chi)\left\|\frac{\hat{Z}_n^*-Z_n^*}{\sqrt{n}}\right\|\left\|\frac{\hat{Z}_n^*}{\sqrt{n}}\right\|=O_p(1)o_p(1)[\tilde{C}_0\hat{\wp}_0^2(p)]^{\frac{1}{2}}=o_p(1),$$

The first inequality is from the Cauchy-Schwarz inequality, the second inequality holds because $(\hat{Z}_n^*-Z_n^*)'(\hat{Z}_n^*-Z_n^*)$ and $\hat{Z}_n^{*\prime}\hat{Z}_n^*$ are non-negative definite and the spectral radius theorem, the second to last equality is by Lemma 1 and the compact parameter space assumption (Assumption 4(*ii*)). The proof of $|\mathscr{A}_{12}|=o_p(1)$ follow by similar fashion. For term $\mathscr{A}_{14}$, note that

$$\mathscr{A}_{14}=\frac{\chi'(\hat{Z}_n^*-Z_n^*)'\hat{X}_n^*}{n}\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\frac{\hat{X}_n^{*\prime}\hat{Z}_n\chi}{n},$$

Then,

$$|\mathscr{A}_{14}|\leq O_p(1)o_p(1)\tilde{C}_0\hat{\wp}_0^2(k)O_p(1)[\tilde{C}_0\hat{\wp}_0^2(p)]^{\frac{1}{2}}=o_p(1).$$

As a result, we have $|\mathscr{A}_1|=o_p(1)$ by the triangular inequality.

The second component $\mathscr{A}_2$ is

$$\begin{aligned}
\mathscr{A}_2 &= -\frac{1}{n}\left(\sum_{i=1}^n\hat{x}_i^{*\prime}(P_x^{-1}-I_k)\hat{x}_i^*-\sum_{i=1}^n x_i^{*\prime}(P_x^{-1}-I_k)x_i^*\right)-\frac{1}{n}\left(\sum_{i=1}^n\hat{z}_i^{*\prime}(\Xi^{-1}-I_p)\hat{z}_i^*-\sum_{i=1}^n z_i^{*\prime}(\Xi^{-1}-I_p)z_i^*\right)\\
&= -\mathscr{A}_{21}-\mathscr{A}_{22}
\end{aligned}$$

For term $\mathscr{A}_{21}$, note that

$$\begin{aligned}
\mathscr{A}_{21} &= \frac{1}{n}\sum_{i=1}^n(\hat{x}_i^*-x_i^*)'P_x^{-1}\hat{x}_i^*+\frac{1}{n}\sum_{i=1}^n x_i^{*\prime}P_x^{-1}(\hat{x}_i^*-x_i^*)-\frac{1}{n}\sum_{i=1}^n(\hat{x}_i^*-x_i^*)'\hat{x}_i^*-\frac{1}{n}\sum_{i=1}^n x_i^{*\prime}(\hat{x}_i^*-x_i^*)\\
&= \mathscr{A}_{211}+\mathscr{A}_{212}-\mathscr{A}_{213}-\mathscr{A}_{214}
\end{aligned}$$

and that

$$\begin{aligned}
|\mathscr{A}_{211}| &\leq \frac{1}{n}\sum_{i=1}^n\left[(\hat{x}_i^*-x_i^*)'(\hat{x}_i^*-x_i^*)\right]^{\frac{1}{2}}[\hat{x}_i^{*\prime}P_x^{-1}\hat{x}_i^*]^{\frac{1}{2}}\leq\frac{1}{n}\sum_{i=1}^n\left[(\hat{x}_i^*-x_i^*)'(\hat{x}_i^*-x_i^*)\right]^{\frac{1}{2}}[\hat{x}_i^{*\prime}\hat{x}_i^*]^{\frac{1}{2}}\|P_x^{-1}\|_o\\
&\leq\frac{1}{n}\sum_{i=1}^n\sum_{\tau=1}^k\sup_{x_{\tau,i}}|\hat{x}_{\tau,i}^*-x_{\tau,i}^*|\sup_{x_{\tau,i}}|x_{\tau,i}^*|\cdot\|P_x^{-1}\|_o\leq\frac{1}{n}\sum_{i=1}^n\sum_{\tau=1}^k\sqrt{\frac{\log\log n}{n^{1-a_\tau}}}\left(\sqrt{2a_\tau\log n}+4\sqrt{\frac{\log\log n}{n^{1-a_\tau}}}\right)\|P_x^{-1}\|_o\\
&= o_p(1)O_p(1)O_p(1)=o_p(1),
\end{aligned}$$

where the first inequality comes from the Cauchy-Schwarz inequality, the second inequality holds because P_x^{-1} is non-negative definite, the fourth inequality is by Proposition 1 and the proof of Lemma A.4 in Liu et al. (2012), the second to last equality is due to the compact parameter space assumption. Similarly, we can show $|\mathscr{A}_{212}|, |\mathscr{A}_{213}|, |\mathscr{A}_{214}|= o_p(1)$. Then $|\mathscr{A}_{21}|= o_p(1)$ by the triangular inequality. The desired result of term $|\mathscr{A}_{22}|= o_p(1)$ follow by similar argument used for term $|\mathscr{A}_{21}|$. As a result, $|\mathscr{A}_2|= o_p(1)$.

The third component $|\mathscr{A}_3|$ is

$$\begin{aligned}
\mathscr{A}_3 &= \frac{2}{n}\left(\sum_{i=1}^{n} \hat{z}_i^{*\prime}\Xi^{-1}(\hat{\Gamma}'\hat{x}_i^*) - \sum_{i=1}^{n} z_i^{*\prime}\Xi^{-1}(\Gamma' x_i^*)\right) - \frac{1}{n}\left((\hat{\Gamma}'\hat{x}_i^*)'\Xi^{-1}(\hat{\Gamma}'\hat{x}_i^*) - (\Gamma' x_i^*)'\Xi^{-1}(\Gamma' x_i^*)\right) \\
&= \frac{2}{n}\sum_{i=1}^{n}(\hat{z}_i^* - z_i^*)'\Xi^{-1}(\hat{\Gamma}'\hat{x}_i^*) + \frac{2}{n}\sum_{i=1}^{n} z_i^{*\prime}\Xi^{-1}[(\hat{\Gamma}-\Gamma)'\hat{x}_i^*] + \frac{2}{n}\sum_{i=1}^{n} z_i^{*\prime}\Xi^{-1}[\Gamma'(\hat{x}_i^* - x_i^*)] \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}[(\hat{\Gamma}-\Gamma)'\hat{x}_i^*]'\Xi^{-1}(\hat{\Gamma}'\hat{x}_i^*) - \frac{1}{n}\sum_{i=1}^{n}[\Gamma'(\hat{x}_i^* - x_i^*)]'\Xi^{-1}(\hat{\Gamma}'\hat{x}_i^*) - \frac{1}{n}\sum_{i=1}^{n}(\Gamma' x_i^*)'\Xi^{-1}[(\hat{\Gamma}-\Gamma)'\hat{x}_i^*] \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}(\Gamma' x_i^*)'\Xi^{-1}[\Gamma'(\hat{x}_i^* - x_i^*)]
\end{aligned}.$$

The three key terms are the ones associated with $\hat{\Gamma}-\Gamma$, the results that other terms are all $o_p(1)$ follow by similar argument for term $|\mathscr{A}_{211}|$. We show the result for the second term, by the second stage estimation,

$$\begin{aligned}
\|\hat{\Gamma}-\Gamma\|_o &\le \left\|\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\right\|_o \left\|\frac{\hat{X}_n^{*\prime}U_n}{n}\right\| + \left\|\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\right\|_o \left\|\frac{\hat{X}_n^*}{\sqrt{n}}\right\| \left\|\frac{\hat{Z}_n^* - Z_n^*}{\sqrt{n}}\right\| \\
&\quad + \left\|\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\right\|_o \left\|\frac{\hat{X}_n^*}{\sqrt{n}}\right\| \left\|\frac{(\hat{X}_n^* - X_n^*)\Gamma}{\sqrt{n}}\right\| \\
&\le O_p(1)o_p(1) + O_p(1)\left[\tilde{C}_0\hat{\wp}_0^2(k)\right]^{\frac{1}{2}} o_p(1) + O_p(1)\left[\tilde{C}_0\hat{\wp}_0^2(k)\right]^{\frac{1}{2}} o_p(1) = o_p(1),
\end{aligned}$$

therefore,

$$\begin{aligned}
\left|\frac{2}{n}\sum_{i=1}^{n} z_i^{*\prime}\Xi^{-1}[(\hat{\Gamma}-\Gamma)'\hat{x}_i^*]\right| &\le \frac{2}{n}\sum_{i=1}^{n}(z_i^{*\prime}\Xi'^{-1}\Xi^{-1}z_i^*)^{\frac{1}{2}}\left(\hat{x}_i^{*\prime}(\hat{\Gamma}-\Gamma)(\hat{\Gamma}-\Gamma)'\hat{x}_i^*\right)^{\frac{1}{2}} \\
&\le \frac{2}{n}\sum_{i=1}^{n}(z_i^{*\prime}z_i^*)^{\frac{1}{2}}(\hat{x}_i^{*\prime}\hat{x}_i^*)^{\frac{1}{2}}\|\Xi^{-1}\|_o\|\hat{\Gamma}-\Gamma\|_o \\
&= O_p(1)O_p(1)O_p(1)o_p(1) = o_p(1),
\end{aligned}$$

then $|\mathscr{A}_3|= o_p(1)$ and we have that

$$\frac{1}{n}\left[\ln L_n(\theta_{ML}) - \ln L_{n0}(\theta_{ML})\right] = o_p(1).$$

The required results for the differences in the first and second order derivatives, i.e., $\frac{1}{n}\left[\frac{\partial \ln L_n(\theta_{ML})}{\partial\theta_l} - \frac{\partial \ln L_{n0}(\theta_{ML})}{\partial\theta_l}\right] = o_p(1)$ and $\frac{1}{n}\left[\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\theta_{l_1}\partial\theta_{l_2}} - \frac{\partial^2 \ln L_{n0}(\theta_{ML})}{\partial\theta_{l_1}\partial\theta_{l_2}}\right] = o_p(1)$, follow by similar fashion and we omit the proof.

Proof of Lemma 3. Let $\theta_{ML,0} = \left(\omega_0', \sigma_{\xi,0}^2, \chi_0', \alpha_0', \delta_0'\right)'$ be the true parameter vector, and $\theta_{ML} = \left(\omega', \sigma_\xi^2, \chi', \alpha', \delta'\right)'$ be arbitrary parameter values in Θ defined in Assumption 4(*ii*). Since $\ln x \le x - 1$

for any $x \geq 0$, which implies $\ln x \leq 2\sqrt{x} - 2$ for any $x \geq 0$, we have

$$
\begin{aligned}
&\frac{1}{n}\mathrm{E}\ln[L_n(\theta_{ML})/L_n(\theta_{ML,0})] \\
&\quad\leq \frac{2}{n}\mathrm{E}(\sqrt{L_n(\theta_{ML})/L_n(\theta_{ML,0})} - 1) = \frac{2}{n}\int\left(\sqrt{L_n(\theta_{ML})/L_n(\theta_{LM,0})} - 1\right)L_n(\theta_{ML,0})\,d\mathscr{U}_n \\
&\quad= \frac{2}{n}\int\sqrt{L_n(\theta_{ML})L_n(\theta_{ML,0})}d\mathscr{U}_n - \frac{1}{n} = -\frac{1}{n}\int\left[\sqrt{L_n(\theta_{ML})} - \sqrt{L_n(\theta_{ML,0})}\right]^2 d\mathscr{U}_n \leq 0
\end{aligned}
$$

where $\mathscr{U}_n = (Y_n', \mathrm{vec}(Z_n)', \mathrm{vec}(X_n)')'$. This implies in particular the information inequality that $\frac{1}{n}\mathrm{E}\ln L_n(\theta_{ML}) \leq \frac{1}{n}\mathrm{E}\ln L_n(\theta_{ML,0})$ for all θ. Thus $\theta_{ML,0}$ is a maximizer. Also, this inequality implies that if $\frac{1}{n}\mathrm{E}\ln L_n(\theta_{ML}) = \frac{1}{n}\mathrm{E}\ln L_n(\theta_{ML,0})$, we have $\frac{1}{n}\ln L_n(\theta_{ML}) = \frac{1}{n}\ln L_n(\theta_{ML,0})$ almost surely. By Lemma 2, this is equivalent to $\frac{1}{n}\ln L_{n0}(\theta_{ML}) + o_p(1) = \frac{1}{n}\ln L_{n0}(\theta_{ML,0}) + o_p(1)$ almost surely, i.e.,

$$
\begin{aligned}
&-\frac{n}{2}\ln\sigma_\xi^2 - \frac{n}{2}\ln|P_x| - \frac{n}{2}\ln|\Xi| + \ln|I_n - \Lambda(\zeta, Z_n)W_n| - \frac{1}{2}\sum_{i=1}^n (z_i^* - \Gamma' x_i^*)'\Xi^{-1}(z_i^* - \Gamma' x_i^*) - \frac{1}{2}\sum_{i=1}^n x_i^{*\prime} P_x^{-1} x_i^* \\
&-\frac{1}{2\sigma_\xi^2}\left\{[I_n - \Lambda(\zeta, Z_n)W_n]Y_n - X_n\beta - (\mathscr{O}_n^\perp Z_n^*)\chi\right\}'\left\{[I_n - \Lambda(\zeta, Z_n)W_n]Y_n - X_n\beta - (\mathscr{O}_n^\perp Z_n^*)\chi\right\} \\
= &-\frac{n}{2}\ln\sigma_{\xi,0}^2 - \frac{n}{2}\ln|P_{x,0}| - \frac{n}{2}\ln|\Xi_0| + \ln|I_n - \Lambda(\zeta_0, Z_n)W_n| - \frac{1}{2}\sum_{i=1}^n (z_i^* - \Gamma_0' x_i^*)'\Xi_0^{-1}(z_i^* - \Gamma_0' x_i^*) - \frac{1}{2}\sum_{i=1}^n x_i^{*\prime} P_{x,0}^{-1} x_i^* \\
&-\frac{1}{2\sigma_{\xi,0}^2}\left\{[I_n - \Lambda(\zeta_0, Z_n)W_n]Y_n - X_n\beta_0 - (\mathscr{O}_n^\perp Z_n^*)\chi_0\right\}'\left\{[I_n - \Lambda(\zeta_0, Z_n)W_n]Y_n - X_n\beta_0 - (\mathscr{O}_n^\perp Z_n^*)\chi_0\right\}
\end{aligned}
\tag{A.3}
$$

holds for Y_n, Z_n, and X_n almost surely. In this proof, we focus on the endogenous heterogeneity specification, the identification for the endogenous W_n setting follows by similar argumenr and is provided in Lemma S.1 in the supplement file.

Differentiate Eq.(A.3) with respect to Y_n, we have

$$
\begin{aligned}
&\sigma_\xi^{-2}[I_n - \Lambda(\zeta, Z_n)W_n]'\left\{[I_n - \Lambda(\zeta, Z_n)W_n]Y_n - X_n\beta - (\mathscr{O}_n^\perp Z_n^*)\chi\right\} \\
&= \sigma_{\xi,0}^{-2}[I_n - \Lambda(\zeta_0, Z_n)W_n]'\left\{[I_n - \Lambda(\zeta_0, Z_n)W_n]Y_n - X_n\beta_0 - (\mathscr{O}_n^\perp Z_n^*)\chi_0\right\}.
\end{aligned}
\tag{A.4}
$$

Differentiate Eq.(A.4) with respect to Y_n once more,

$$
\sigma_\xi^{-2}[I_n - \Lambda(\zeta, Z_n)W_n]'[I_n - \Lambda(\zeta, Z_n)W_n] = \sigma_{\xi,0}^{-2}[I_n - \Lambda(\zeta_0, Z_n)W_n]'[I_n - \Lambda(\zeta_0, Z_n)W_n], \tag{A.5}
$$

where $\Lambda(\zeta, Z_n) \equiv \mathrm{diag}\{\lambda(\zeta, z_1), ..., \lambda(\zeta, z_n)\}$, because of the stochasticity of Z_n, from Eq.(A.5), $1/\sigma_\xi^2 = 1/\sigma_{\xi,0}^2$. Hence, $\sigma_\xi = \sigma_{\xi,0}$ and we should have $[I_n - \Lambda(\zeta, Z_n)W_n]'[I_n - \Lambda(\zeta, Z_n)W_n] = [I_n - \Lambda(\zeta_0, Z_n)W_n]'[I_n - \Lambda(\zeta_0, Z_n)W_n]$ almost surely, which implies

$$
\begin{aligned}
&\{\mathrm{diag}[\lambda(\zeta_0, z_1) - \lambda(\zeta, z_1), ..., \lambda(\zeta_0, z_n) - \lambda(\zeta, z_n)]\}W_n + W_n'\{\mathrm{diag}[\lambda(\zeta_0, z_1) - \lambda(\zeta, z_1), ..., \lambda(\zeta_0, z_n) - \lambda(\zeta, z_n)]\} \\
&- W_n'\left\{\mathrm{diag}\left[\lambda^2(\zeta_0, z_1) - \lambda^2(\zeta, z_1), ..., \lambda^2(\zeta_0, z_n) - \lambda^2(\zeta, z_n)\right]\right\}W_n = 0.
\end{aligned}
$$

As Z_n is stochastic, it must be that $\lambda(\zeta, z_i) = \lambda(\zeta_0, z_i)$ holds almost sure for each i. By Assumption 4(*iii*), it indicates that $\zeta = \zeta_0$. Then Eq.(A.4) implies

$$
[I_n - \Lambda(\zeta_0, Z_n)W_n]'X_n(\beta - \beta_0) + [I_n - \Lambda(\zeta_0, Z_n)W_n]'(\mathscr{O}_n^\perp Z_n^*)(\chi - \chi_0) = 0,
$$

As $[I_n - \Lambda(\zeta_0, Z_n)W_n]$ is invertible, X_n and $\mathscr{O}_n^\perp Z_n^*$ are not linearly dependent, we must have $X_n(\beta -$

$\beta_0) = 0$ and $(\mathscr{O}_n^{\perp} Z_n^*)(\chi - \chi_0) = 0$. Therefore, $\beta = \beta_0$ and $\chi = \chi_0$. Eq.(A.3) implies that

$$\begin{aligned} &-\frac{n}{2}\ln|P_x| - \frac{n}{2}\ln|\Xi| - \frac{1}{2}\sum_{i=1}^{n}(z_i^* - \Gamma' x_i^*)'\Xi^{-1}(z_i^* - \Gamma' x_i^*) - \frac{1}{2}\sum_{i=1}^{n} x_i^{*\prime} P_x^{-1} x_i^* \\ &= -\frac{n}{2}\ln|P_{x,0}| - \frac{n}{2}\ln|\Xi_0| - \frac{1}{2}\sum_{i=1}^{n}(z_i^* - \Gamma_0' x_i^*)'\Xi_0^{-1}(z_i^* - \Gamma_0' x_i^*) - \frac{1}{2}\sum_{i=1}^{n} x_i^{*\prime} P_{x,0}^{-1} x_i^*, \end{aligned} \tag{A.6}$$

the above equation only involves z_j^* and x_j^* for $j = 1, ..., n$, which are strictly increasing functions of z_j and x_j. First, by differentiating Eq.(A.6) with respect to z_j^*, we have

$$\left(z_j^* - \Gamma' x_j^*\right)' \Xi^{-1} = \left(z_j^* - \Gamma_0' x_j^*\right)' \Xi_0^{-1} \tag{A.7}$$

Differentiate Eq.(A.7) once more with respect to z_j^*, we can get $\Xi^{-1} = \Xi_0^{-1}$. By the uniqueness of a matrix inverse, $\Xi = \Xi_0$. Then Eq.(A.7) implies $(\Gamma - \Gamma_0)' x_j^* \Xi_0^{-1} = 0$, which indicates that $\Gamma = \Gamma_0$ as x_j^* is stochastic.[41] Second, given $\Xi = \Xi_0$ and $\Gamma = \Gamma_0$, differentiating Eq.(A.7) with respect to x_i^* twice yields $P_x^{-1} = P_{x,0}^{-1}$, by the uniqueness of a matrix inverse, $P_x = P_{x,0}$. Thus, we have $\theta_{ML} = \theta_{ML,0}$, $\theta_{ML,0}$ is the unique maximizer of $\ln L_n(\theta_{ML,0})$.

Proof of Theorem 1. For a SAR model with endogenous W_n, all terms in the log pseudo-likelihood function in Appendix A.3.1 can be written in the general formula $(\theta_{ML,0} - \theta_{ML})' \frac{1}{n} a \boldsymbol{\varphi}_n^{*\prime} \mathcal{M}_n \boldsymbol{\varphi}_n^* b(\theta_{ML,0} - \theta_{ML})$, where a, b are some constant vectors, $\boldsymbol{\varphi}_n^* = (\varphi_1^*, ..., \varphi_n^*)'$ with $\varphi_i^* = f_i(v_i, X_n, Z_n, \theta_{ML,0})$ being a vector value function, $\mathcal{M}_n = A_n' B_n$ are either $W_n^{m_1}$ or $G_n^{m_2}$ with m_1 and m_2 being finite non-negative integers. Therefore, with Assumption 7, the proof of consistency and asymptotic normality follow by similar arguments for the proof of Theorem 3 in Qu and Lee (2015) as direct applications of Proposition 1, Corollary 1 and Proposition 2 in their paper.

For a SAR model with endogenous heterogeneity, with the identification condition in Assumption 7, first we check the consistency of the 3SPMLE in two steps. In the first step, we consider the uniform convergence $\frac{1}{n}\left[\sup_{\theta_{ML}\in\Theta} |\ln L_n(\theta_{ML}) - \mathrm{E}\ln L_n(\theta_{ML})|\right] \xrightarrow{p} 0$. By Lemma S.2 in the supplement file, we have the result that $\sup_{\zeta\in\Theta_\zeta} \frac{1}{n}\left(\ln|I_n - \Lambda(\zeta, Z_n) W_n| - \mathrm{E}\ln|I_n - \Lambda(\zeta, Z_n) W_n|\right) \xrightarrow{p} 0$, it remains to check the uniform convergence of the remaining terms in the log pseudo-likelihood function in Appendix A.3.1, which can be expressed in the general formula $a' \boldsymbol{\varphi}_n^{*\prime} \tilde{\mathcal{M}}_n \boldsymbol{\varphi}_n^* b$, where a and b are some constant vectors, $\tilde{\mathcal{M}}_n = \tilde{A}_n' \tilde{B}_n$, $\tilde{A}_n$ and $\tilde{B}_n$ are either $\mathscr{D}_n [\Lambda(\zeta, Z_n) W_n]^{m_1}$ or $\tilde{G}_n^{m_2}(\zeta)$ (or $\mathscr{D}_n [\Lambda(\zeta_0, Z_n) W_n]^{m_1}$ or $\tilde{G}_n^{m_2}(\zeta_0)$) with m_1 and m_2 being finite non-negative integers, $\mathscr{D}_n = \mathrm{diag}(\boldsymbol{d}(\zeta, z_i))$ is a diagonal matrix of some globally bounded function of ζ and z. Hence, by Proposition S.1(i) in the supplement file, we have the pointwise convergence of the aforementioned terms. As $\frac{1}{n}[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp} \hat{Z}_n^*)\chi]'[V_n(\omega) - (\hat{\mathscr{O}}_n^{\perp} \hat{Z}_n^*)\chi]$ in Appendix A.3.1 can be expressed as the general formula $\frac{1}{n} a' \varphi_n^{*\prime} \tilde{G}_n^{m_1}(\zeta)' \tilde{G}_n^{m_2}(\zeta) \varphi_n^* b$ where $\tilde{G}_n(\zeta) = \mathscr{D}_n W_n [I_n - \Lambda(\zeta, Z_n) W_n]^{-1}$, the desired stochastic equicontinuity result follows from Proposition S.1(ii) in the supplement file. In the second step, we consider the uniform equicontinuity of $\lim_{n\to\infty} \mathrm{E}\left(\frac{1}{n}\ln L_n(\theta_{ML})\right)$. With the results that $\frac{1}{n}\mathrm{E}\left|a' \boldsymbol{\varphi}_n^{*\prime} \tilde{\mathcal{M}}_n \boldsymbol{\varphi}_n^* b\right| = O(1)$ (Proposition S.1 (i) in the supplement file) and the boundedness of the parameter space (Assumption 4*(ii)*), by inequality in (S.30) of the supplement file, we have the required result. Then $\hat{\theta}_{ML} \xrightarrow{p} \theta_{ML,0}$. Next, we the asymptotic normality of $\hat{\theta}_{ML}$. We can write the second derivatives in Appendix A.3.2 as the general formula in Proposition S.1(ii) in the supplement file, then we have the uniform convergence $\frac{1}{n}\sup_{\theta_{ML}\in\Theta}\left\|\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\theta\partial\theta'} - \mathrm{E}\left(\frac{\partial^2 \ln L_n(\theta_{ML})}{\partial\theta\partial\theta'}\right)\right\| \xrightarrow{p} 0$. Applying the CLT in Proposition S.1(iii) in the supplement file to $\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_{ML,0})}{\partial\theta_{ML}}$ in Appendix A.3.2, because

41. Γ can also be identified from the second stage estimation, $\|\Gamma - \Gamma_0\|_o = o_p(1)$ as shown in the proof of Lemma 2.

$\mathrm{E}\left(\frac{\partial \ln L_n(\theta_{ML,0})}{\partial \theta_{ML}} \frac{\partial \ln L_n(\theta_{ML,0})}{\partial \theta'_{ML}}\right) = -\mathrm{E}\left(\frac{\partial^2 \ln L_n(\theta_{ML,0})}{\partial \theta_{ML} \partial \theta'_{ML}}\right)$ by Assumptions 1 and 2, we have

$$
\begin{aligned}
\sqrt{n}(\hat{\theta}_{ML} - \theta_{ML,0}) &= -\left(\frac{\partial^2 \ln L_n(\bar{\theta}_{ML})}{\partial \theta_{ML} \partial \theta'_{ML}}\right)^{-1} \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_{ML,0})}{\partial \theta_{ML}} \\
&\xrightarrow{d} \mathrm{N}\left(0, \left(\lim_{n\to\infty} \frac{1}{n}\mathrm{E}\left(\frac{\partial^2 \ln L_n(\theta_{ML,0})}{\partial \theta_{ML} \partial \theta'_{ML}}\right)\right)^{-1} \lim_{n\to\infty} \frac{1}{n}\mathrm{E}\left(\frac{\partial \ln L_n(\theta_{ML,0})}{\partial \theta_{ML}} \frac{\partial \ln L_n(\theta_{ML,0})}{\partial \theta'_{ML}}\right)\right. \\
&\qquad \left. \times \left(\lim_{n\to\infty} \frac{1}{n}\mathrm{E}\left(\frac{\partial^2 \ln L_n(\theta_{ML,0})}{\partial \theta_{ML} \partial \theta'_{ML}}\right)\right)^{-1}\right) \\
&\xrightarrow{d} \mathrm{N}\left(0, -\left(\lim_{n\to\infty} \frac{1}{n}\mathrm{E}\left(\frac{\partial^2 \ln L_n(\theta_{ML,0})}{\partial \theta_{ML} \partial \theta'_{ML}}\right)\right)^{-1}\right).
\end{aligned}
$$

Proof of Theorem 2. We can express

$$
\begin{aligned}
\hat{\theta}_{IV,\cdot} - \theta_{IV,\cdot,0} &= [\hat{M}'_{n,\cdot}\hat{T}_{n,\cdot}(\hat{T}'_{n,\cdot}\hat{T}_{n,\cdot})^{-1}\hat{T}'_{n,\cdot}\hat{M}_{n,\cdot}]^{-1}\hat{M}'_{n,\cdot}\hat{T}_{n,\cdot}(\hat{T}'_{n,\cdot}\hat{T}_{n,\cdot})^{-1}\hat{T}'_{n,\cdot}\hat{\boldsymbol{\epsilon}}_{n,\cdot} \\
&= [\frac{\hat{M}'_{n,\cdot}\hat{T}_{n,\cdot}}{n}(\frac{\hat{T}'_{n,\cdot}\hat{T}_{n,\cdot}}{n})^{-1}\frac{\hat{T}'_{n,\cdot}\hat{M}_{n,\cdot}}{n}]^{-1}\frac{\hat{M}'_{n,\cdot}\hat{T}_{n,\cdot}}{n}(\frac{\hat{T}'_{n,\cdot}\hat{T}_{n,\cdot}}{n})^{-1}\left(\frac{1}{n}\hat{T}'_{n,\cdot}\boldsymbol{\epsilon}_n + \frac{1}{n}\hat{T}'_{n,\cdot}(\hat{\mathscr{O}}_n U_n \gamma_{\cdot,0})\right. \\
&\quad \left. + \frac{1}{n}\hat{T}'_{n,\cdot}[\hat{\mathscr{O}}_n^{\perp}(Z_n^* - \hat{Z}_n^*)\gamma_{\cdot,0}] - \frac{1}{n}\hat{T}'_{n,\cdot}[\hat{\mathscr{O}}_n^{\perp}(X_n^* - \hat{X}_n^*)\Gamma_{\cdot,0}\gamma_{\cdot,0}]\right).
\end{aligned}
$$

Note that $\hat{M}_{n,\cdot} = M_{n,\cdot} + \left(0, \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*\right)$ and $\hat{T}_{n,\cdot} = T_{n,\cdot} + \left(0, \hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*\right)$, the asymptotic analysis relies on terms of

$$
\begin{aligned}
&\frac{1}{n}T'_{n,\cdot}\boldsymbol{\epsilon}_{n,\cdot},\ \frac{1}{n}T'_{n,\cdot}(\hat{\mathscr{O}}_n U_n \gamma_{\cdot,0}),\ \frac{1}{n}T'_{n,\cdot}[\hat{\mathscr{O}}_n^{\perp}(Z_n^* - \hat{Z}_n^*)\gamma_{\cdot,0}],\ \frac{1}{n}T'_{n,\cdot}[\hat{\mathscr{O}}_n^{\perp}(X_n^* - \hat{X}_n^*)\Gamma_{\cdot,0}\gamma_{\cdot,0}],\ \frac{1}{n}T'_n(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*) \\
&\frac{1}{n}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*)'\boldsymbol{\epsilon}_{n,\cdot},\ \frac{1}{n}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*)'(\hat{\mathscr{O}}_n U_n \gamma_{\cdot,0}),\ \frac{1}{n}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*)'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*), \\
&\frac{1}{n}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*)'[\hat{\mathscr{O}}_n^{\perp}(Z_n^* - \hat{Z}_n^*)\gamma_{\cdot,0}],\ \frac{1}{n}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*)'[\hat{\mathscr{O}}_n^{\perp}(X_n^* - \hat{X}_n^*)\Gamma_{\cdot,0}\gamma_{\cdot,0}].
\end{aligned}
\tag{A.8}
$$

For term $\frac{1}{n}T'_{n,\cdot}\boldsymbol{\epsilon}_{n,\cdot}$, denote $t_{i,w} = \left(q'_{i,w}, x'_i, \mathscr{O}'_i\right)$ and $t_{i,h} = \left(q'_{1i,h}, ..., q'_{pi,h}, x'_i, \mathscr{O}'_i\right)$ with $\mathscr{O}_i = \sum_{j=1}^n \hat{\mathscr{O}}_{ij}^{\perp}\hat{z}_j^{*\prime}$, the key is to consider $q_{i,w} = \sum_{j=1}^n w_{ij}x_j$, $q_{\iota i,h} = \lambda_\iota(z_{\iota i})\sum_{j=1}^n w_{ij}x_j$ for $\iota = 1, ..., p$ because other terms in $t_{i,\cdot}$ satisfy required moment properties by assumption. For all i, $\sum_{j\neq i} c_1 d_{ij}^{-c_0 d_0} < \infty$ by Claim C.1.1 in Qu and Lee (2015), and $|\lambda_\iota(z_{\iota i})| < \infty$ under Assumption 4(*iii*), we have

$$
\begin{aligned}
\mathrm{E}(|q_{\iota i,h}q'_{\iota i,h}|) &= \mathrm{E}(|\lambda_\iota^2(z_{\iota i})\sum_{j_1=1}^n w_{ij_1}x_{j_1}\sum_{j_2=1}^n w_{ij_2}x_{j_2}|) \le \mathrm{E}(|\lambda_\iota^2(z_{\iota i})|\sum_{j_1=1}^n\sum_{j_2=1}^n |w_{ij_1}||w_{ij_2}||x_{j_1}x'_{j_2}|) \\
&\le c_2 \sum_{j_1\neq i} c_1 d_{ij_1}^{-c_0 d_0} \sum_{j_2\neq i} c_1 d_{ij_2}^{-c_0 d_0} \sup_{n.j_1,n.j_2}\mathrm{E}|x_{j_1}x'_{j_2}| < \infty,
\end{aligned}
$$

similarly $\mathrm{E}(|q_{i,w}q'_{i,w}|) < \infty$, then $\mathrm{Var}\left(\frac{1}{n}T'_{n,\cdot}\boldsymbol{\epsilon}_{n,\cdot}\right) = \frac{\eta_{\cdot,0}^2}{n^2}\mathrm{E}(T'_n T_n) \le \frac{\eta_{\cdot,0}^2}{n}\sup_{n.i}\mathrm{E}(t_i t'_i) = O(\frac{1}{n})$. Therefore, $\frac{1}{n}T'_{n,\cdot}\boldsymbol{\epsilon}_{n,\cdot} = O_p(\frac{1}{\sqrt{n}})$. For term $\frac{1}{n}T'_{n,\cdot}(\hat{\mathscr{O}}_n U_n \gamma_{\cdot,0})$, as

$$
\|\frac{1}{n}T'_{n,\cdot}(\hat{\mathscr{O}}_n U_n \gamma_{\cdot,0})\|_o \le \left\|\frac{T_{n,\cdot}}{\sqrt{n}}\right\| \left\|\frac{\hat{X}_n^*}{\sqrt{n}}\right\| \left\|\left(\frac{\hat{X}_n^{*\prime}\hat{X}_n^*}{n}\right)^{-1}\right\|_o \left\|\frac{\hat{X}_n^{*\prime}U_n\gamma_{\cdot,0}}{n}\right\| \le O_p(1)\left[\tilde{C}_0\wp_0^2(k)\right]^{\frac{1}{2}} O_p(1)o_p(1) = o_p(1),
$$

then $\frac{1}{n}T'_{n,\cdot}(\hat{\mathscr{O}}_n U_n \gamma_{\cdot,0}) = o_p(1)$. By the result of Lemma 1 and the proof in Proposition 2, it can be easily shown that all the remaining terms related to $(\hat{X}_n^* - X_n^*)$, $(\hat{Z}_n^* - Z_n^*)$, and $(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^* - \mathscr{O}_n^\perp Z_n^*)$ in (A.8) are $o_p(1)$. As a result, the first and second stage estimation errors will not affect the asymptotic analysis of the 3rd stage, and we can treat $\hat{X}_n^*$, $\hat{Z}_n^*$ and $\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*$ as the true X_n^*, Z_n^* and $\mathscr{O}_n^\perp Z_n^*$. Under Assumptions 1-5, 8(*i*)-8(*ii*) and 9, the asymptotic properties can be obtained by applying Proposition 1 in Qu and Lee (2015) and Proposition S.1 in the supplement file. Besides, we can apply a related CLT (Proposition 2 in Qu and Lee (2015) for the endogenous W_n specification and Proposition S.1(iii) for the endogenous heterogeneity setting) to get the results that $\frac{1}{\sqrt{n}}T'_{n,\cdot}\boldsymbol{\epsilon}_{n,\cdot} = \frac{1}{\sqrt{n}}\sum_{i=1}^n t'_{i,\cdot}\epsilon_{i,\cdot} \xrightarrow{d} \mathrm{N}(0, \eta^2_{\cdot,0}T'_{n,\cdot}T_{n,\cdot})$ and that $\frac{1}{\sqrt{n}}T'_{n,\cdot}(\hat{\mathscr{O}}_n U_n \gamma_{\cdot,0}) \xrightarrow{d} \mathrm{N}\left(0, T'_{n,\cdot}(\gamma'_{\cdot,0}\Xi_{\cdot,0}, \gamma_{\cdot,0}\hat{\mathscr{O}}_n)T_{n,\cdot}\right)$ where $\Xi_0 = P_{z,0} - P'_{xz,0}P_{x,0}^{-1}P_{xz,0}$, then we have the desired asymptotic distribution.

Proof of Lemma 4. By Eq.(24) and (25), denote $G_n = S_n^{-1}W_n$ with $S_n = I_n - \lambda_0 W_n$, and $\tilde{\tilde{G}}_{\iota,n} = \Lambda_\iota(z_\iota)W_n S_n^{-1}$ with $S_n = I_n - \sum_{\iota=1}^p \varrho_{\iota,0}\Lambda_\iota(z_\iota)W_n$, we have

$$\begin{aligned}\tilde{\boldsymbol{\epsilon}}_{n,w} &= S_n(\hat{\lambda})Y_n - X_n\hat{\beta} - \hat{\mathscr{O}}_n^\perp \hat{Z}_n^*\hat{\gamma} = S_n(\hat{\lambda})S_n^{-1}(X_n\beta_0 + \hat{\mathscr{O}}_n^\perp \hat{Z}_n^*\gamma_0 + \hat{\boldsymbol{\epsilon}}_{n,w}) \\ &= (\lambda_0 - \hat{\lambda})G_n(X_n\beta_0 + \hat{\mathscr{O}}_n^\perp \hat{Z}_n^*\gamma_0) + X_n(\beta_0 - \hat{\beta}) + \hat{\mathscr{O}}_n^\perp \hat{Z}_n^*(\gamma_0 - \hat{\gamma}) + (\lambda_0 - \hat{\lambda})G_n\hat{\boldsymbol{\epsilon}}_{n,w} + \hat{\boldsymbol{\epsilon}}_{n,w},\end{aligned}$$

and

$$\begin{aligned}\tilde{\boldsymbol{\epsilon}}_{n,h} &= S_n(\hat{\zeta})Y_n - X_n\hat{\beta} - \hat{\mathscr{O}}_n^\perp \hat{Z}_n^*\hat{\gamma} = S_n(\hat{\zeta})S_n^{-1}(X_n\beta_0 + \hat{\mathscr{O}}_n^\perp \hat{Z}_n^*\gamma_0 + \hat{\boldsymbol{\epsilon}}_{n,h}) \\ &= \sum_{\iota=1}^p(\varrho_{\iota,0} - \hat{\varrho}_\iota)\tilde{\tilde{G}}_{\iota,n}(X_n\beta_0 + \hat{\mathscr{O}}_n^\perp \hat{Z}_n^*\gamma_0) + X_n(\beta_0 - \hat{\beta}) + \hat{\mathscr{O}}_n^\perp \hat{Z}_n^*(\gamma_0 - \hat{\gamma}) + \sum_{\iota=1}^p(\varrho_{\iota,0} - \hat{\varrho}_\iota)\tilde{\tilde{G}}_{\iota,n}\hat{\boldsymbol{\epsilon}}_{n,h} + \hat{\boldsymbol{\epsilon}}_{n,h}.\end{aligned}$$

where the composite error $\hat{\boldsymbol{\epsilon}}_{n,\cdot} = \boldsymbol{\epsilon}_{n,\cdot} + \hat{\mathscr{O}}_n U_n\gamma_{\cdot,0} + \hat{\mathscr{O}}_n^\perp(Z_n^* - \hat{Z}_n^*)\gamma_{\cdot,0} - \hat{\mathscr{O}}_n^\perp(X_n^* - \hat{X}_n^*)\Gamma_{\cdot,0}\gamma_{\cdot,0}$. We can express $\frac{1}{n}\tilde{\boldsymbol{\epsilon}}'_{n,w}\tilde{\boldsymbol{\epsilon}}_{n,w}$ in the form below as in Proposition 1 in Qu and Lee (2015) so that:

$$\begin{aligned}\frac{1}{n}\tilde{\boldsymbol{\epsilon}}'_{n,w}\tilde{\boldsymbol{\epsilon}}_{n,w} &= \left(\theta_{IV,w,0} - \hat{\theta}_{IV,w}\right)'\frac{1}{n}a_1\varphi_n^{*\prime}\mathscr{M}_n\boldsymbol{\varphi}_n^* b_1\left(\theta_{IV,w,0} - \hat{\theta}_{IV,w}\right) + \frac{1}{n}a_2\boldsymbol{\varphi}_n^{*\prime}\mathscr{M}_n\boldsymbol{\varphi}_n^* b_2\left(\theta_{IV,w,0} - \hat{\theta}_{IV,w}\right) \\ &\quad + \frac{1}{n}\hat{\boldsymbol{\epsilon}}'_{n,w}\hat{\boldsymbol{\epsilon}}_{n,w} \xrightarrow{p} \eta^2_{w,0}\end{aligned}$$

Besides, $\frac{1}{n}\tilde{\boldsymbol{\epsilon}}'_{n,h}\tilde{\boldsymbol{\epsilon}}_{n,h}$ can be expressed in the following form as in Proposition S.1(i) in the supplement file so that:

$$\begin{aligned}\frac{1}{n}\tilde{\boldsymbol{\epsilon}}'_{n,h}\tilde{\boldsymbol{\epsilon}}_{n,h} &= \left(\theta_{IV,h,0} - \hat{\theta}_{IV,h}\right)'\frac{1}{n}a_3\varphi_n^{*\prime}\tilde{\mathscr{M}}_n\boldsymbol{\varphi}_n^* b_3\left(\theta_{IV,w,0} - \hat{\theta}_{IV,w}\right) + \frac{1}{n}a_4\boldsymbol{\varphi}_n^{*\prime}\tilde{\mathscr{M}}_n\boldsymbol{\varphi}_n^* b_4\left(\theta_{IV,h,0} - \hat{\theta}_{IV,h}\right) \\ &\quad + \frac{1}{n}\hat{\boldsymbol{\epsilon}}'_{n,h}\hat{\boldsymbol{\epsilon}}_{n,h} \xrightarrow{p} \eta^2_{h,0}\end{aligned}$$

where $\tilde{\mathscr{M}}_n = \tilde{A}'_n\tilde{B}_n$, and $\tilde{A}_n$ and $\tilde{B}_n$ are combinations of $[\Lambda_{\iota_1}(z_{\iota_1})W_n]^{m_\iota}$ $(\iota = 1, ..., p)$.[42] Therefore, $\frac{1}{n}\tilde{\boldsymbol{\epsilon}}'_{n,\cdot}\tilde{\boldsymbol{\epsilon}}_{n,\cdot} \xrightarrow{p} \eta^2_{\cdot,0}$.

To show $\hat{\mathcal{U}}_{IV,w} \xrightarrow{p} \mathcal{U}_{IV,w}$ and $\hat{\mathcal{U}}_{IV,h} \xrightarrow{p} \mathcal{U}_{IV,h}$, the most complicated terms to be analyzed are

$$\begin{aligned}&\frac{1}{n}\left[a'_5(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)'G_n(\hat{\lambda})'G_n(\hat{\lambda})(\hat{\mathscr{O}}_n^\perp Z_n^*)b_5 - \mathrm{E}\left(a'_5(\mathscr{O}_n^\perp Z_n^*)'G'_nG_n(\mathscr{O}_n^\perp Z_n^*)b_5\right)\right] = o_p(1), \text{ and} \\ &\frac{1}{n}\left[a'_6(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)'\tilde{\tilde{G}}_{\iota,n}(\hat{\zeta})'\tilde{\tilde{G}}_{\iota,n}(\hat{\zeta})(\hat{\mathscr{O}}_n^\perp \hat{Z}_n^*)b_6 - \mathrm{E}\left(a'_6(\mathscr{O}_n^\perp Z_n^*)'\tilde{\tilde{G}}'_{\iota,n}\tilde{\tilde{G}}_{\iota,n}(\mathscr{O}_n^\perp Z_n^*)b_6\right)\right] = o_p(1), \ \iota = 1, ..., p.\end{aligned} \tag{A.9}$$

42. Because $\Lambda_{\iota_1}(z_{\iota_1})W_n$ is a special case of $\Lambda(\zeta, Z_n)W_n$ and $[I_n - \sum_{\iota=1}^p \varrho_{\iota,0}\Lambda_\iota(z_\iota)W_n]^{-1} = \sum_{k=1}^\infty[\sum_{\iota=1}^p \varrho_{\iota,0}\Lambda_\iota(z_\iota)W_n]^k$ with $[\sum_{\iota=1}^p \varrho_{\iota,0}\Lambda_\iota(z_\iota)W_n]^k = \sum_{i_1+i_2+\cdots+i_p=k; i_1,i_2,\ldots,i_p\geq 0}\begin{pmatrix} k \\ i_1, i_2, \ldots, i_p\end{pmatrix}\cdot\prod_{\iota=1}^p[\varrho_{\iota,0}\Lambda_\iota(z_\iota)W_n]^{i_\iota}$, the NED properties preserve under multiplication for the component $\prod_{\iota=1}^p[\varrho_{\iota,0}\Lambda_\iota(z_\iota)W_n]^{i_\iota}$ by Claim B.3 in Qu and Lee (2015). It's easy to obtain the similar LLN as in Proposition S.1(i) in the supplement file.

Similar to the proof in Proposition 2, it can be verified that $\frac{1}{n}(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^* - \mathscr{O}_n^{\perp}Z_n^*) = o_p(1)$, then we have

$$\frac{1}{n}[a_5'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'G_n(\hat{\lambda})'G_n(\hat{\lambda})(\hat{\mathscr{O}}_n^{\perp}Z_n^*)b_5 - \mathrm{E}(a_5'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'G_n(\hat{\lambda})'G_n(\hat{\lambda})(\hat{\mathscr{O}}_n^{\perp}Z_n^*)b_5)] = o_p(1)$$

from the ULLN (Corollary 1) in Qu and Lee (2015). By the ULLN in Proposition S.1(ii) in the supplement file,

$$\frac{1}{n}[a_6'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'\tilde{\tilde{G}}_{\iota,n}(\hat{\zeta})'\tilde{\tilde{G}}_{\iota,n}(\hat{\zeta})(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)b_6 - \mathrm{E}(a_6'(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)'\tilde{\tilde{G}}_{\iota,n}(\hat{\zeta})'\tilde{\tilde{G}}_{\iota,n}(\hat{\zeta})(\hat{\mathscr{O}}_n^{\perp}\hat{Z}_n^*)b_6)] = o_p(1), \ \iota = 1, ..., p.$$

Terms in (A.9) hold from the equicontinuity of $\frac{1}{n}\mathrm{E}[a_5'U_n'(\Gamma)G_n(\lambda)'G_n(\lambda)U_n(\Gamma)b_5]$ and $\frac{1}{n}\mathrm{E}[a_6'U_n'(\Gamma)\tilde{\tilde{G}}_{\iota,n}(\zeta)'\tilde{\tilde{G}}_{\iota,n}(\zeta)U_n(\Gamma)b_6]$.

References

Anselin, L. 1988. *Spatial Econometrics: Methods and Models.* The Netherlands: Kluwer Academic.

Balakrishnan, N., and C. Lai. 2009. *Continuous bivariate distributions.* New York: Springer.

Barbe, P., C. Genest, K. Ghoudi, and R. Rémillard. 1996. "On kendall's process." *Journal of Multivariate Analysis* 58:197–229.

Becker, J., D. Proksch, and C. Ringle. 2021. "Revisiting Gaussian copulas to handle endogenous regressors." *Journal of the Academy of Marketing Science* 50:1–21.

Case, A., H. Rosen, and J. Hines. 1993. *Budget spillovers and fiscal policy interdepence: evidence from the states. In: Giles, D., Ullah, A. (Eds.), Handbook of Applied Economics Statistics.* New York: Marcel Dekker.

Cliff, A., and J. Ord. 1973. *Spatial Autocorrelation.* London: Pion Ltd.

Danaher, P., and M. Smith. 2011. "Modeling multivariate distributions using copulas: Applications in marketing." *Marketing Science* 30 (1): 4–21.

Deheuvels, P. 1979. "La fonction de dépendance empirique et ses propriétés - Un test non paramétrique d'indépendance." *Académie Royale de Belgique - Bulletin de la Classe des Sciences - 5e Série* 65:274–292.

Genest, C., and L. Rivest. 1993. "Statistical inference procedures for bivariate Archimedean copulas." *Journal of the American Statistical Association* 88 (423): 1034–1043.

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.

Gupta, A., and P. Robinson. 2015. "Inference on higher-order spatial autoregressive models with increasingly many parameters." *Journal of Econometrics* 186:19–31.

Guth, D., and S. Zhang. 2022. "Geographic Spillover Effects of Prescription Drug Monitoring Programs (PDMPs)." Working Paper, https://arxiv.org/pdf/2107.04925.pdf.

Han, F., T. Zhao, and H. Liu. 2013. "CODA: high dimensional copula discriminant analysis." *Journal of Machine Learning Research* 14:629–671.

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.

He, M., Y. Chen, and R. Schramm. 2018. "Technological spillovers in space and firm productivity: Evidence from China's electric apparatus industry." *Urban Studies* 55 (11): 2522–2541.

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

Kelejian, H., and I. Prucha. 2001. "On the asymptotic distribution of the Moran I test statistic with applications." *Journal of Econometrics* 104:219–257.

Lee, L. 2004. "Asymptitic distributions of quasi-maximum likelihood estimators for spatial econometric models." *Econometrica* 72:1899–1925.

———. 2007. "GMM and 2SLS estimation of mixed regressive, spatial autoregressive models." *Journal of Econometrics* 137:489–514.

Li, X., P. Mikusiński, H. Sherwood, and M. Taylor. 1997. *On approximation of copulas, in V. Benes and J. Stěpán (Eds.), Distributions with Given Marginals and Moment Probelems.* Dordrecht: Kluwer Academic Publishers.

Lin, Y., Y. Song, and Y. Yang. 2022. "QML estimation of a spatial autoregressive model with endogenous heterogeneity." Working Paper, .

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.

Malikow, E., and Y. Sun. 2017. "Semiparametric estimation and testing of smooth coefficient spatial autoregressive models." *Journal of Econometrics* 199 (1): 12–34.

Nelsen, R. 2006. *An Introduction to copulas.* New York: Springer.

Ord, J. 1975. "Estimation methods for models of spatial interaction." *Journal of the American Statistical Association* 70 (349): 120–126.

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

Pinkse, J., and M. Slade. 2010. "The future of spatial econometrics." *Journal of regional Science* 50 (1): 103–117.

Qu, X., and L. F. Lee. 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.

Rabovič, R., and P. Čížek. 2022. "Estimation of spatial sample selection models: A partial maximum likelihood approach." Journal of Econometrics (*Forthcoming*), https://doi.org/10.1016/j.jeconom.2021.10.011, January.

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

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.

Song, P. 2000. "Multivariate dispersion models generated from gaussian copula." *Scandinavian Journal of Statistics* 27 (2): 305–320.

Stoyanov, A., and N. Zubanov. 2012. "Productivity Spillovers across Firms through Worker Mobility." *American Economic Journal: Applied Economics* 4 (2): 168–198.

Su, L., and S. Jin. 2010. "Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models." *Journal of Econometrics* 157 (1): 18–33.

———. 2018. "Sieve maximum likelihood estimation of the spatial autoregressive Tobit model." *Journal of Econometrics* 203 (1): 96–112.

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.

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.